

**Asgeir Drøivoldsmo**

**New tools and technology for the study of  
human performance in simulator  
experiments**

**Dr.polit.- thesis 2003**

**Dept. of Psychology**

**Faculty of Social Sciences and Technology Management**

**Norwegian University of Science and Technology,**

**NTNU Trondheim**

**ISBN Number 82-471-5235-5**

## ACKNOWLEDGEMENTS

The project team with members from the virtual reality (VR), reactor operation, radiation protection and psychology divisions at the Halden Reactor Project participated in the development and planning of the experiments. The same group has participated through the specification, design, prototyping and evaluation phases in four experiments. I would like to express my gratitude to the project team members, in particular Thomas Elisenberg, Geir Mjønes, and Heidi Andersen who made this project possible with their co-operative attitude, broad operational experience and insightful help during the whole project. Niels-Kristian Mark, who helped build the VR-model of the reactor hall, and Terje Johnsen who did the programming part of the human performance data collection system presented as a part of these thesis. This system was used for collecting the performance data in the experiments. Special thanks to Stein Helgar who has done a considerable contribution in both VR modelling and data collection. He has also helped creating a number of graphics and the videos prepared for this thesis. Finally, I would like to thank the operators and the other staff at the OECD Halden Reactor project for participating as subjects in the four experiments, and last but not least Ivar A. Bjørgen (NTNU) and Jan Heim (SINTEF) for using their invaluable knowledge and great experience pointing me in the right direction both during the experiments and finalising the writing of this thesis.

## ABSTRACT

This thesis suggests that new tools and technology can be used for production of relevant data and insights from the study of human performance in simulator and field experiments. It examines some of the theoretical perspectives behind data collection and human performance assessment, and argues for a high resemblance of the real world and use of subject matter expertise in simulator studies. A model is proposed, suggesting that human performance measurement should be tightly coupled to the topic of study and have a close connection to the time line. This coupling requires new techniques for continuous data collection, and eye movement tracking has been identified as a promising basis for this type of measures.

One way of improving realism is to create virtual environments allowing for controlling more of the environment surrounding the test subjects. New application areas for virtual environments are discussed for use in control room and field studies.

The combination of wearable computing, virtual and augmented (the use of computers to overlay virtual information onto the real world) reality provides many new possibilities to present information to operators. In two experiments, virtual and augmented reality techniques were used to visualise radiation fields for operators in a contaminated nuclear environment. This way the operators could train for and execute their tasks in a way that minimised radiation exposure to the individual operator. Both experiments were successful in proving the concept of radiation visualisation

Virtual environments allow for early end-user feedback in the design and refurbishment of control room man-machine interfaces. The practical usability of VR in the control room setting was tested in two control room design experiments. The results show that with the right tools for solving the tasks under test, even desktop presentations of the virtual environment can provide sufficient resemblance of the real world.

Computerised data collection was identified as a key factor for efficient data handling in simulator and field experiments. An Internet based questionnaire system was specified and implemented in parallel with the experimental work, and functionality was developed to fulfil the special needs of each new experimental setting. The result is a computer platform independent system capable of handling most types of data and test situations for data collection by Internet or local computer networks.

## Table of Contents

1	OUTLINE AND OBJECTIVE.....	1
1.1	Scope.....	1
1.2	Experimental work and tool development.....	1
1.3	Outline.....	1
1.4	Objective .....	2
2	INTRODUCTION .....	5
2.1	Engineering psychology.....	5
2.2	Investigating human performance.....	6
2.2.1	The scientific approach.....	6
2.2.2	Applied experiments, maximising the use of expensive data.....	7
2.3	Synthetic worlds.....	8
2.4	Wearable computing.....	10
3	PROCESS CONTROL SYSTEMS.....	13
3.1	Operation.....	14
3.2	Modernising the control room.....	16
3.3	Human factors in control room design.....	17
4	SIMULATOR STUDIES.....	22
4.1	Simulation.....	22
4.2	Simulators.....	22
4.3	Extending the simulation.....	23
4.3.1	HAMMLAB as an example.....	23
4.3.2	Control room example.....	24
4.3.3	Field operator example.....	25
4.4	Simulators as a way of gathering knowledge about human problem-solving.....	26
4.4.1	Validity.....	27
4.4.2	Ecological validity and simulation.....	28
4.4.3	Experimental control.....	29
4.4.4	Close to everything can be simulated, but not yet - an example...	33
5	HUMAN PERFORMANCE ASSESSMENT IN SIMULATION.....	34
5.1	Complex problem-solving.....	34
5.2	Selecting indicators of human performance.....	37

5.2.1	The need for process expertise.....	39
5.3	Data collection methods and measures overview.....	40
5.3.1	Measuring operator cognitive problem-solving.....	41
5.3.2	Questionnaire based data collection using the Internet.....	43
5.3.3	Operator control actions.....	45
5.3.4	Measuring the plant performance.....	45
5.4	Measurement problems.....	45
5.4.1	Instructional effects on self-reporting.....	45
5.4.2	Strengths and weaknesses of the SME.....	46
5.5	Data scoring and analysis.....	48
5.5.1	Process tracing methods.....	48
6	DISCUSSION OF THE EXTENDED SIMULATION.....	51
6.1	Limitations of the extended simulation.....	51
6.1.1	SOP definition.....	51
6.1.2	Factors affecting SOP.....	52
6.2	Extended control room simulation, artificial case example.....	52
6.2.1	Upgrading hybrid control rooms.....	53
6.2.2	A VR control environment for upgrading NPP turbine systems.....	53
6.3	VR as a tool for simulation of field operations in control room studies. .	55
6.4	Final comments on psychological experiments in VR.....	57
7	CONCLUSION.....	60
8	REFERENCES.....	62
8.1	List of acronyms and abbreviations.....	70

Paper A  
Paper B  
Paper C  
Paper D  
Paper E  
Paper F  
Paper G  
Paper H  
Paper I

## Papers and reports

The papers are either submitted to scientific journals (1), book-chapters (2), presented on international conferences (3) or reports for the OECD Halden Reactor Project member countries (4). The data collection tool is entirely Internet based and can be tested by accessing <http://psyweb.svt.ntnu.no/asgeird/> and following the instructions for registration.

- A) Drøivoldsmo, A. and Johnsen, T. (2002). Development and testing of an Internet-based data collection technique for simulator and real world experimentation. Manuscript submitted for publication. (1)
- B) Drøivoldsmo, A. & Louka, M.N. (2002) Virtual reality tools for testing control room concepts. In B.G. Lipták (Ed.), *Instrument Engineers' Handbook, Vol.3* (pp. 299-306). Boca Raton, FL: CRC PRESS LLC. (2)
- C) Drøivoldsmo, A., Reigstad, M., Shimoda, H., Louka, M.N., Helgar, S., Gustavsen, M. & Nystad, E. (2002). *Augmented reality for visualisation of radiation (HWR-717)*. Halden, Norway: OECD Halden Reactor Project. (3, 4) Videos supporting the method section of this paper is located at <http://psyweb.svt.ntnu.no/asgeird>
- D) Drøivoldsmo, A., Nystad, E. & Helgar, S. (2001). *Virtual reality verification of workplace design guidelines for the process plant control room (HWR-633)*. Halden, Norway: OECD Halden Reactor Project. (3, 4)
- E) Drøivoldsmo, A. (2000) Study of errors by means of simulation and training. In P.F. Elzer, R.H. Kluwe & B. Boussoffara (Eds.), *Human error and system design and management* (pp. 201-206). London: Springer-Verlag. (3)
- F) Drøivoldsmo, A., Skraaning, G., Sverrbo, M., Dalen, J., Grimstad, T. & Andresen, G. (1998). *Continuous measures of situation awareness and workload (HWR-539)*. Halden, Norway: OECD Halden Reactor Project. (3, 4)
- G) Nystad, E., Drøivoldsmo, A. & Sebok, A. (2002). Use of radiation maps in a virtual training environment for NPP field operators. *Proceedings of the man-machine systems research sessions at the enlarged Halden programme group meeting, Gol, Norway, 8th to 13 September 2002 (volume 2.)* pp. C5 P1. Halden, Norway: OECD Halden Reactor Project. (3, 4)
- H) Nystad, E., Helgar, S. & Drøivoldsmo, A. (2002). Virtual reality verification of workplace design guidelines: A follow-up study. *Proceedings of the man-machine systems research sessions at the enlarged Halden programme group meeting, Gol, Norway, 8th to 13 September 2002 (volume 2.)* pp. C5 P9. Halden, Norway: OECD Halden Reactor Project. (3, 4)
- I) Drøivoldsmo, A. (2002). *The Halden Questionnaire help pages [On-line]*. Available: <http://psyweb.svt.ntnu.no/asgeird/>: Internet.





# **1 OUTLINE AND OBJECTIVE**

## ***1.1 Scope***

This document and a number of papers and a data collection tool constitute the thesis. The main objective of the work done for this thesis is to produce new knowledge, tools and ideas for how to effectively use the experimental paradigm in assessment of human performance pertaining to the area of engineering psychology. Based on the needs of engineering psychology in design and operation of human - machine systems, operation of the nuclear power plant (NPP) has been used as the framework for the discussion and practical work. Both the control room and the field operator's situation in control and operation of the power plant have been used for discussion, tool development and basis for forming the experimental hypothesis tested. Based upon literature, interviews, site visits, virtual reality (VR) control room experiments and field operation experiments, future changes in the way of running simulator experiments and performing data collection has been identified. The discussion is based in the research needs in the nuclear power industry, with focus on the need for good indicators for measuring human performance and the advantages and problems by introduction of new technology.

## ***1.2 Experimental work and tool development***

Four experiments have been carried out with the purpose of evaluating real world application and usability of VR and augmented reality (AR) technology. Two of the experiments have been carried out with the specific purpose of testing the practical usability of VR in the control room setting, while the other two have been in the area of VR and augmented reality for field operation. The results from the experiments are part of the foundation for the discussion. The insights produced are forming a basis for the examples of future application of the VR and AR technology in computerised simulation with a high degree of ecological validity.

As an answer to the need of more efficient handling of data from the type of experiments performed and discussed, the area of computerised data collection was identified as a key factor for improvement. An Internet based questionnaire system was specified and implemented in parallel with the experimental work and functionality was developed to fulfil the special needs of each new experimental setting. The result is a computer platform independent system capable of handling most types of data and test situations for data collection by Internet or local computer networks.

## ***1.3 Outline***

Starting with the methodological foundation for the work and introduction of the domain for application, the need and methods for future human performance assessment in control room design for the advanced process control simulators are discussed. Experimental work in the Halden Man-Machine LABORatory (HAMMLAB) and the

Halden Virtual Reality Centre (HVRC) is used as basis for the discussion. Examples of how the new technology can be implemented in simulators and simulator studies are provided to give the reader better insight in the problem area. Further, the question about ecological validity and an attempt to develop an eclectic model of human performance has been given a central position in the discussion. Methods and measures are discussed with the purpose of developing a concept of new designs and extended simulations, taking advantage of new technology in the form of wearable computing and the techniques of VR and AR. Finally, the question of limitations and future possibilities is discussed and exemplified by an artificial case bringing synthetic worlds into simulator based control room experimentation.

## **1.4 Objective**

*The accident at Three Mile Island 2 began at 4 a. m. on March 28. A minor malfunction, or transient, in the nonnuclear part of the system would evolve a series of automated responses in the reactor's coolant system, and during all of this, the relief valve on top of a piece of equipment called the "pressurizer" would become stuck open. Owing to continued misreading of the symptoms by the operators over a 2 ¼-hour period before the relief valve was closed and the turning off of an automatic emergency cooling system, the reactor core would become partially uncovered and severely damaged. It would be another 12 hours before the plant crew and the engineers from GPU Service Company would concur in effective corrective action. And that was but the beginning.*

The citation above is from Rogovin and Framptons (1980) report to the commissioners and the public. Three Mile Island (TMI) is a sad example of the fact that inadequate design of control systems is a major source for erroneous human actions and remains a main threat to the safe operation of high-risk systems. On the other hand, the events in the early phase of the TMI accident in the spring of 1979 are probably among the events that have put most focus on the human operator in the process control system and triggered the most human centred research in the design of safe systems for the last two decades. In an effort to address this design question, Moray and Huey, (1988) list the following four central aspects of human performance, for which it is necessary to obtain more knowledge.

1. Be able to measure human performance in existing systems.
2. Understand and predict the effects of changes in human performance of the overall human-technical system and the effects of changes in human performance that would result from proposed modifications to the technical system.
3. Predict human performance in situations that are by design expected to occur rarely, if ever, and which cannot be tested in the operating system.
4. Predict human performance in new systems especially before they are operational, indeed very early in the conceptual design stages.

The OECD Halden Reactor Project has for many years delivered insights from an experimental programme pertaining to this area. Applying simulators and close to real world studies of human performance in the HAMMLAB and the HVRC, the issues of

human performance and design have been in focus. One basic assumption underlying this research has been that the interconnectedness in complex systems gives rise to important phenomena, which can be missed studying simpler subsystems separately (Drøivoldsmo, 2000; Skraaning, 2002). A main problem in the study of realistic situations is to maintain the control with the variables influencing performance; e.g., the ability to preserve the experimental control. In NPP simulation studies there are multiple solutions to almost any problem and the situation is dynamically changing based on the operator's control actions. The number of possible problem solutions and the operator's opportunity to choose among different strategies with an almost equal final outcome raises the question whether it is possible to measure the effect of an experimental manipulation? If such measurement is possible, how should the measures be designed and applied to distinguish good and bad human performance? Furthermore, the results from human performance measurement need to be transformed into design process input based on a human centred approach. In this respect, VR is a promising powerful and cost effective tool in design work. It enables the design staff to identify and eliminate potential problems early in the design process, and enables end-users to actively participate in the design process. The technology has for some time been used for visualisation and training personnel for different tasks, but due to the high need of computer performance, the area has been almost entirely driven by the technology, and the lack of real world applications has been noticeable. Therefore, more knowledge needs to be obtained about the quality of results from utilisation of VR technology and particularly in the area of human performance measurement in synthetic worlds.

To address the questions stated above, the specific objectives of this thesis are as follows:

- Provide a model for realistic experimentation and human performance measurement in dynamic high complexity situations.

- Establish and discuss the criteria for use of new front edge technology like VR in human performance experimentation.

- Provide examples of how the combination of the two preceding objectives can benefit the human interface design process.



## 2 INTRODUCTION

### 2.1 *Engineering psychology*

"It is the discipline of engineering psychology, specifically, that applies a psychological perspective to the problems of system design" (Wickens and Hollands, 2000, p. 2). They further state, that among the notable features of engineering psychology as it has emerged as a discipline in the last four decades, are its solid theoretical basis and its close relationship to the study of experimental psychology, but engineering psychology differs from the experimental psychology with respect to the requirement for experiment design. The goal of experimental psychology is to uncover laws of behaviour, but there is not, as it is in engineering psychology, a constraining requirement to apply the laws.

To define engineering psychology it is also necessary to give a short introduction to the human factors and ergonomics area. Human factors is traditionally thought of as the profession that studies the human-machine interface, and has its roots in the US military problems during and after World War II. The term 'ergonomics' is much older and originates from the introduction of machinery in the industrial revolution. Ergonomics is linked to industrial applications and uses information from work physiology, biomechanics and anthropometry for design of workstations and industrial processes (Hellander, 1997). Human factors can be said to extend the concept of ergonomics with a higher emphasis on the cognitive and social aspects involved in human machine interaction. Cacciabue (1998) defines human factors as the professional speciality "concerned with the analysis and optimisation of the relationship between people and their activities by the integration of human sciences and system engineering in systematic applications and working environment frameworks".

The main purpose of human factors and ergonomics is design. This is thereby different from the areas of anthropology, cognitive science, psychology, sociology, and medical sciences, since their primary purpose is to understand and model human behaviour, but not to utilise the knowledge for design (Hellander, 1997). Wickens and Hollands (2000) states the fundamental goal of human factors in engineering is to reduce error, increase productivity, and enhance safety and comfort when humans interact with a system. Human factors as a discipline involves the study of factors and development of tools that help to achieve these goals. They further describe the area of human factors as a broad area concerning much more than the information-processing capacities of the brain, which falls inside the realm of psychology.

Engineering psychology differs from human factors in scope. It is an applied research component of disciplines such as human factors and ergonomics. The role of engineering psychology is to provide information about human performance capabilities (Howell, 1993). A typical human factors project would be to compare the relevance of two different designs for equipment, while the engineering psychology task would be to specify the capacities and limitations of the human, provide experimental data for this purpose, and in this way make a basis available for making a selection of the best design. "Although research topics in engineering psychology are selected because of applied needs, the research goes beyond specific one-time applications and has the broader objective of providing a usable theory of human performance" (Wickens and

Hollands, 2000, p. 3). Engineering psychology has often been referred to as applied experimental psychology because of its close links to experimental psychology.

## **2.2 Investigating human performance**

There are several paths to the goal of building knowledge for prediction of human performance in the area of process control systems design. Transfer of new knowledge from the established sciences listed above, is of course one of the main sources. But the insights from basic scientific research need to be adapted to the applied settings, and there is an additional need for more specific research in the domain of engineering psychology. A number of methods are available and they are more or less appropriate depending on the task to be investigated (task analysis, incident reports, surveys, experiments, simulations etc.). Relevance for the purpose, cost and time consumption are often the factors deciding the choice of methodology. There is probably no single solution to this choice, but design relevance is one of the most critical qualities to achieve. Furthermore it is reasonable to assume that more realistic circumstances of evaluation lead to greater confidence that the results of the methodology will generalise to the product in question (Wickens and Hollands, 2000), but realistic circumstances are not sufficient alone. To produce the theory needed for prediction of human performance in human-machine systems there is a need for a scientific approach, implying careful and systematic interpretation of the phenomenon under investigation. The fact that the theory is applied in safety critical domains puts even stronger requirements on the quality assurance of the scientific work.

### **2.2.1 The scientific approach**

To collect the data and build the theories valuable for improving human performance measurement we need to be able to distinguish science from pseudoscience. Karl Popper proposed what is known as falsification. "A theory is 'scientific' if one is prepared to specify in advance a crucial experiment (or observation) which can falsify it." Even if Popper's "quick kill by refutation" or what Lakatos (1978) calls "naive falsification" today not is accepted as a sufficient criteria for demarcation between science and pseudoscience, it has set the standard for what we know as modern science. Popper's theory is characterised by logically involving predictions of possible theories. Following this, the theories can be falsified. This forms the basis for what we know as the hypothetical – deductive method. The most structured way to test the hypothesis is to do experiments. The necessity of experiments for production of design relevant data is further substantiated by the following:

*To acquire facts relevant for the identification and specification of the various processes at work in nature it is, in general, necessary to practically intervene to try to isolate the processes under investigation and eliminate the effects of others. In short, it is necessary to do experiments (Chalmers, 1999, p.28).*

Going as far as discussing hypothesis and intervention, raises the question about time and cost as factors for deciding the methodology. Unfortunately, deciding the choice of methodology for the questions under investigation involves being under constant pressure from the second critical factor mentioned above. Cost and time will especially

come into consideration in an applied area when the motivation for starting a project is mostly economical. The results from designing an acceptable Human - Machine Interface (HMI) cannot always be measured on a short-term basis, and hence the investments in the human operators is reduced, preferring the more measurable parts of the system.

### **2.2.2 Applied experiments, maximising the use of expensive data**

In applied experiments for the industry the falsification approach is not fruitful alone. The need for data and insights from full scale simulation in almost any area of human process control, leads to the conclusion that the methods and tests leading to more and better insights into a problem area should be chosen. Stringent testing and purely hypothetical - deductive experimentation should not be regarded as the only fruitful way of scientific development in HMI design.

To produce knowledge for the domain of human factors the experiment should be used where applicable, but with the case of limited knowledge and resources, it should not be a requirement for all production of valuable data as input for the profession of human factors specialists. There are situations where the hypothesis should be tested in an experiment, like in the final validation of a design, as a test of whether the proposed design fulfils the requirements for safe operation, and for validation, major design inventions, methods and measures. On the other hand, with the limited resources available for the production of high quality data and observations, one should seek to get the most out of every costly data collection. One commonly used technique for this is to apply the quasi-experimental method. The quasi-experiments considered in this document are for the re-use of a simulations database. The problem with quasi-experimentation is that since subjects are not randomly assigned to the independent variables, one is not assured against undetected nuisance variables. Undetected nuisance variables are a threat to the internal validity of a study because of the risk that variables other than the independent variable are responsible for variation in the dependent variable (Kirk, 1995). However, through careful control of variables and use of designs, more threats to internal validity can be ruled out (Cook and Campbell, 1979).

To maximise the outcome of the quasi-experimental approach, expensive full-scale simulations could often give a better outcome with a conservative use of conditions in the experimental test, e.g., instead of tweaking the maximum number of conditions out of every experiment, a conservative estimate of subjects in every cell could be used to improve the situation for post hoc analysis. A higher N in each cell would build a better database for quasi-experimentation and the reuse of data would be more attractive. In addition to the need for a sufficient N, there is also a need for a standardised test. One way of solving this problem would be to develop baseline scenarios for a small number of relevant operating situations and use these scenarios as an extensible basis for all experiments.

Other known techniques try to eliminate as many variables as possible without unnecessary data collection. The procedure for multifactor studies proposed by Simon (1973; cited in Meister, 1985) and sequential experimentation (Han, Williges and Williges, 1997) are sophisticated techniques trying to grasp more of the complexity of the systems under evaluation. The basic procedure is to find as many potential

independent variables as possible and then to use a stepwise procedure reducing the number of variables actually influencing the experiment. However, as objected by Meister (1985), most of the potential variables in an experimental design can be excluded by careful expert evaluation, and the complexity and cost of such procedures make them inaccessible for most researchers. A cost-effective way of narrowing down the number of variables could be to adapt to the procedure described by Rouse (1991, chap. 7) using multiple methods like expert paper evaluations and part-task simulations before going to the full scale simulations.

## **2.3 Synthetic worlds**

In the future design process for human-machine systems the use of new technology could play a central role. With the techniques of VR and AR computers can be used to represent synthetic replacements of systems and areas not easily accessed in the real world, but a number of parameters need to be settled before successful application of this technology can take place.

A definition of VR will have to include the system's ability to represent a physical environment as well as the interactive aspects; the user should be able to navigate in the virtual space or manipulate virtual objects. We will adopt the definition used by Wann and Mon-Williams (1996), where a virtual environment (VE) is described as an environment with the following properties:

*A virtual environment (VE) provides the user with access to information that would not otherwise be available at that place or time, capitalizes upon natural aspects of human perception by extending visual information in three spatial dimensions and may supplement this information with other sensory stimuli and temporal changes. ... a further constraint to introduce is that a VE enables the user to interact with the displayed data (p. 833).*

Successful introduction of VR technology for the purpose of control room design and verification demands technologically simple and inexpensive solutions which are manageable without immense investments. Reducing cost by dropping multimodal interaction elements (e.g. tactile sensors, head trackers, etc.) is one way of dealing with this problem. The approach used in the referred studies (Drøivoldsmo et al., 2000, 2002; Drøivoldsmo, Nystad and Helgar, 2001; Nystad, Helgar and Drøivoldsmo, 2002; Nystad, Drøivoldsmo and Sebok, 2002; Sebok, Nystad and Drøivoldsmo, 2002) relies on good visualisation and motion for depth cues, and should be inside of the proposed definition by a safe margin. This way of representing the environment is often referred to as desktop VR (Wickens and Hollands, 2000) or Fish-Tank VR (Heim, 1998).

Azuma (Azuma, 2001) defines an AR system as one that “supplements the real world with virtual (computer-generated) objects that appear to coexist in the same space as the real world”. Wearing an head mounted display (HMD), for example, the user of an AR system can see the real world around him, but with additional virtual objects appropriately aligned (or “registered,” in AR terminology) with real objects. As the user moves around in the real environment the virtual objects maintain correct alignment with the physical environment resulting in a seamless combination of real and virtual objects. AR can thus enhance the user's perception of the real world. When defining



AR, Azuma lists the following three properties that an AR system should typically exhibit:

combines real and virtual objects in a real environment

runs interactively, and in real time; and

registers (aligns) real and virtual objects with each other.

This definition is not restricted to visual information and is equally applicable to the other senses, however, most existing AR systems place an emphasis on the display of visual information. In the mobile AR system used in the experiment described later in this report, visual, auditory, and tactile information are combined.

AR is often viewed by researchers (e.g. Azuma, 1997) as a variation of VR and the two concepts are sometimes confused, however, VR and AR are conceptually very different. While VR technology is used to immerse the user in a VE, replacing the real world with an entirely computer-generated one, AR seeks to supplement the real world, in which the user is already immersed, with additional information. Milgram (Milgram, 1994) describes AR as a mixed reality, placed between the real environment and a VE. Milgram's reality-virtuality continuum is shown in Figure 1. Augmented virtuality is another type of mixed reality, where real objects are added to virtual ones, as opposed to augmented reality where virtual objects are added to real ones.



Figure 1 Milgram's reality-virtuality continuum.

Sutherland laid the foundations of AR research in the 1960s, however, AR is relatively young as a research field in its own right, with the bulk of the research work done during the last decade.

Researchers are pursuing the values and possibilities of AR. The essential value of AR is that it can enhance human perception and extend conventional interaction between man and the environment. Moreover, it has the potential to create a powerful new style of interaction. In a sense, AR can be viewed as an "Intelligence Amplifier", as proposed by Brooks (Brooks, 1996).

If the computerised presentation has insufficient resemblance to a natural situation, the subject's belief in the VE may decline, and the usefulness of the simulation will decrease for a number of purposes.

The assumption of a successful extended simulation rests on the subject's sense of presence (SOP) in the situation simulated. Witmer & Singer (1998) defined SOP as "the subjective experience of being in one place, even when one is physically situated in another", or as "experiencing the computer-generated environment rather than the actual physical locale" (p. 225). Regenbrecht, Schubert & Friedman (1998) emphasise the cognitive construction of a space in which the user's body and the VE seems to share the same world. However, as Stanney & Salvendy (1998) and Wilson (1997) point out, there is an absence of agreed methodology and measures for the SOP in a VE.

For the purpose of the current discussion, SOP in a VE may be divided into two factors. First the environmental factor manifested in the computerised model perceived by the subject entering the VE. Second, by the subject's belief in the usefulness and realism of the VE presented. This division has its roots in the Cartesian view, postulated by the French philosopher Rene Descartes as the dualism between objects located inside or outside the mind. Immanuel Kant also used this dualism, dividing between the object as an object for possible experience and the object in itself (das Ding an sich). Other thinkers like Hegel, Heidegger and Gibson are exponents of alternative views, where the experience of reality is in interaction with the surroundings and that subjective and objective experience cannot be separated from one another due to the dynamic interaction and reciprocity of effects (Haga, 2002, Lübce, 1983).

Sheridan (1999) does attempt to accommodate both views in what he calls an "eclectic ontology of presence" using engineering and an information-processing model based on estimation theory. One important feature of this model is the building stone saying that "Real" reality can never be known but, because of sensory and action constraints, only estimated (p. 556). It is a fact that all understanding of the world is mediated by human perception, and for the use of VR/AR in experimentation the subjective experience of the simulated world is of high importance, but merely to the point of sufficient presence. Estimating a threshold for sufficient subjective presence in a VE would be a good way of ruling out VEs not fulfilling the requirements for an adequate simulation, but it would not be sufficient for rating the quality of the VE. For this purpose objective measurement would be needed.

## **2.4 Wearable computing**

For the last few years, there has been a small revolution in the areas of computer graphics and wireless communication. In the near future, operators will be able to wear a computer much in the same way as eyeglasses or clothing are worn today, and they will be able to interact with the computer based on the context of the situation. With head-up displays, unobtrusive input devices, high-speed wireless local area networks, and context sensing and communication tools, the wearable computer can act as an intelligent assistant to personnel performing work where conventional computers are unsuitable. These continuous technological advancements influence a number of aspects with regard to the human operator.

Skourup and Reigstad (2002) define a wearable computer as "a compact, self-contained portable, full functioning computing device which is completely supported by a user for the input, processing and output of information".

For the control room operator this new technology gives the opportunity to operate without being located in front of a specific workstation. For the field operator, presentation of plant status, alarm information and multimedia communication are examples of areas where wearable computing can potentially prove efficient. For the design of human - machine control systems this opportunity for extended interaction between control room and field operators will have an impact on the allocation of functions as well as the need for information presentation on both ends of the (wireless) line.

*In a world where personal information devices are a reality, ubiquitous computing is under development, and neural implants are a possibility, the subject of analysis has to be the augmented human, we can expect that studies of human-machine interaction and applied cognitive psychology will converge over time (Chignell, Hancock, Takeshita, 1999).*

This "augmentation" of the operator raises a number of questions about how engineering psychology should treat the new situation with respect to simulation and testing of operators inside and outside the control room.

A typical process plant can be exemplified by the NPP. The nuclear power industry is often regarded as the high-end complexity domain because of its advanced security systems (Moray, 1997). Unlike the smelting plant for e.g., steel or aluminium there are several barriers between the operator (the term operator will be used as a generic term including all control room personnel and workers on the plant outside the control room) and the process, making it physically invisible to the human operator. A physical visible plant allows for more manual operation, and the introduction of automation of operations are often less developed in the smelting industry. Petrochemical refineries and offshore installations lie somewhere between the smelting and the NPP with regard to complexity and degree of automation, but for some plants and off-shore installations the petrochemical industry applies computer systems for more of the operation and control.



### 3 PROCESS CONTROL SYSTEMS

Process control is generally described by a high number of interacting variables, where flows of mass and energy are decided by the dynamics of complex physical and chemical reactions. The operators of the process plant are not, like in the action of controlling a vehicle, dependent so much on motor skills but on decision-making, attention, planning and problem-solving. Operators and systems must be able to predict the process reactions to a controlling action. System responses are often slow and the time from when an operational control action has taken place to when the effective response can be seen, can vary from seconds to hours. This time delay between executing a controlling action such as e.g., increasing or decreasing the flow of liquid in a pipe, makes it difficult to know the actual state of the plant.



*Figure 2. Turbine side of the Swedish Forsmark 3, illustrating a hybrid solution where the monitors in front are used for information gathering and the panels in the back are used for both monitoring and operation.*

Introducing the computer in process control has been a big challenge for both the system designers and the operators using the systems. In the old analogue control rooms there is usually a physical line connecting a single sensor out in the process to a single meter in the control room. These control rooms are referred to as conventional control rooms. In a complex process with many signals, this 1:1 relationship between sensors and the display of information requires huge control rooms with thousands of instruments, and the operator will often need to gather information from several instruments and then work out the relation between them to get the answer to a specific problem. Along with the development of computer systems during the last two decades, there has been extensive activity in upgrading the control rooms with computerised monitoring systems, and to a lesser degree, systems for operation. Control rooms with a mix of computerised and analogue systems are referred to as hybrid control rooms. One of the main purposes of the computers is to collocate information and give the operators more accurate overview of the plant status. A typical hybrid control room can be seen in Figure 2.



*Figure 3. Kashiwazaki unit 6, Japanese advanced NPP control room. Most control panels are replaced with large screen information panels and the process is controlled from the operators workstations.*

Distinct from the conventional and hybrid control-rooms, the advanced control room has fully digitalized both monitoring and control (as shown in Figure 3). The advanced control room uses computer graphics to replace the conventional control panels with a combination of graphical monitors and large screen technology. In the nuclear area, just a few reactors have this type of digitalized control rooms today, but the need for upgrades and the advantages of using computers instead of hardwired panels lead to high activity in upgrading the control rooms.

### **3.1 Operation**

The staff running the NPP consists of a range of personnel from nuclear physicists to mechanics and different security departments. These are necessary support functions for the heart in the plant, which is the control room. The NPP control room operators normally work in crews of three operators; a shift supervisor, a reactor operator and a turbine operator. The different persons in the crew have, as indicated by their position titles, different tasks and are specifically trained for their position in the control room. The reactor operator controls all systems connected to the production of energy from the nuclear fuel, while the turbine operator controls the systems utilising the energy for production of electricity.

In a process plant like a NPP the control room operators have tasks like those listed in Table 1. During start-up and shutdown of the reactor and in the situation of deviations from the normal running of the process, the control room crew has a high number of procedures and tasks to perform. On the other hand, when the process is running stably, the operators could be in the situation of doing nothing besides scheduled adjustments

for several months.

*Table 1. Examples of task performed by the control room staff during normal operation and outages (adapted from Moray, 1997).*

### **Normal operation and maintenance**

- Monitor process
- Adjust system to production and safety goals
- Consult operating procedures and other information
- Keep record log of events
- Shift handover
- Introduce long term changes and adjustments to the system making it more efficient
- Training and retraining

### **Disturbances, faults and emergencies**

In addition to the tasks listed under normal operation and outage, the following tasks are typical during deviations from normal operation

- Detect abnormality
- Diagnose cause of faults
- Counteract disturbances
- Consult emergency operating procedures
- Perform emergency shutdown or other control actions to avoid dangerous accidents

In addition to the control room operators the shift crew is constituted of a number of operators working out in the station. A generic term for this group is field operators. This group can cover a range of professions such as electricians, technicians, and engineers, but normally their tasks are to perform controls and manual operations necessary to maintain and run the plant. The typical tasks of the field operator are described in Table 2 below.

The task of running the plant requires a considerable collaboration between the control room operators and the field operators. Normally the field operators are subordinated operators in the control room where the reactor and turbine operators are supervising their respective personnel. As a part of the operating crew, the field operators have their base in the control room and when they are working out in the plant, the normal mode of communication is by radio or telephone. This collaboration is supported by simulator training of field operators for controlling the plant and working as a field operator is usually the starting point in the education to become a certified control room operator.

One important goal of the work presented in this document is, by using VEs, to provide the domain of human factors with methodology for simulation and data collection, taking functions outside of the control room into account in the work with new systems. In particular, the focus is directed to the future need of verification and validation of new technology available for the field operator, and test methodology needed to investigate the effect of introducing this technology into the NPP.

Table 2. Examples of tasks performed by the field operator.

### **Normal operation and maintenance**

- Monitor process, routine checklist of status out in the process
- Assist control room tests of equipment status
- Perform routine maintenance or necessary maintenance requested by the control room
- Shift handover with control of work orders
- Introduce long term changes and adjustments to the system making it more efficient
- Training and retraining

### **Disturbances, faults and emergencies**

In addition to the tasks listed under normal operation and outage, the following tasks are typical during deviations from normal operation

- Assist control room in detection of abnormality
- Consult emergency operating procedures
- Assist control room in performing control actions to avoid dangerous accidents
- Specific emergency tasks related to, e.g. fire protection

## **3.2 Modernising the control room**

Moving from conventional control to advanced computerised systems has its clear advantages, but it is also burdened with a number of problems. Performing standardised tasks, computers are superior to humans and reallocating such tasks to the more reliable automatic systems has been shown effective. The problem is deciding what tasks to reallocate and how to ensure that the operator is informed about what is going on in the plant. Allowing the computers to automatically control more of the plant's functions has reduced the need for human intervention in the process, but not necessarily reduced the need for operator awareness of plants process status. Technology today would allow building completely autonomous plants, but the designer of the plant will by no means be able to predict all potentially dangerous situations. In situations with uncertainty regarding the solution to a problem, the computerised systems are not always the best tool to decide what to do, and the flexibility and creativity of the human operators is the only protection against inefficiency, failure and in worst case, catastrophe.

In design and physical layout, computerised monitoring and control has introduced what is known as the keyhole problem. In conventional control rooms where instruments and controls are hardwired and located on big control panels, spatial location plays an highly important role in building the operator's awareness of the plant status. The same information is available at one location for all operators and all of the information is always accessible. Accessing information through a computer is different. Even if large screen monitors often are used to present key information critical for monitoring the plant, standard video display units (VDU) are used for operation and control restricting the operators view of the process. The main restriction is that the process has to be split down into different systems and subsystems and presented as a small section of what



used to be the entire view of the process. This introduces a different way of navigation. Earlier the operator could walk to the information; now, in the computerised control room the operator has to request the information from the computer. The obvious advantage of using the computer this way is that the information can be presented in a huge number of different combinations depending on the situation or the operational task the operator is about to perform. For specific tasks in surveillance and decision-making, computer - operator support systems (COSS) can be helpful to the operator.

At the OECD Halden Reactor project, transfer of the control room information from mimic displays into computer monitors has been a topic of investigation for about three decades. In the years 1968 to 1972 a computerised control room with graphical monitors providing alarms, trends and system mimic displays was built for controlling the Halden reactor (Netland and Øyen Hol, 1977). Using four graphical monitors, (three for the operator, shown in Figure 4, and one for the shift supervisor) tracker-balls and small function keyboards, all necessary information and control options for running the Halden Reactor were provided for the operator. In 1972 the reactor was controlled for about 6 weeks using this system. The total computer power used for the system was less than a thousandth of what we get from a standard portable computer today.



*Figure 4. The OPCOM system that for six weeks in 1972 controlled the Halden Boiling Water Reactor*

Today, 25 years later, we have only two fully computerised NPP systems with less than 10 operational units around the world. One of the main reasons for this low number is the limited knowledge on efficient information presentation, and the fear of operator loss of control.

This document presents a proposal for how to integrate old and new systems in early phases of the design, making the systems testable by combining simulators VDU-based systems with VR.

### **3.3 Human factors in control room design**

Human factors requirements to the control room design and functionality are now a commonly accepted idea by the designers working with the man - machine systems in refurbishment and planning of new plants. The process control domain today has a number of well-developed programs for how to consider the human factor in new designs (IEC-964, 1989; ISO-11064-1, 2000; NUREG-0711, 2002). The most recognised models have a similarly general approach to the handling of outputs from the human-system interface design verification and validation. They all enhance the importance of early corrections to the design proposals and recommend an iterative process where verification and validation results should be used as input for the next phase in the design process. A summary based on the best known standards is presented in Figure 5 below.

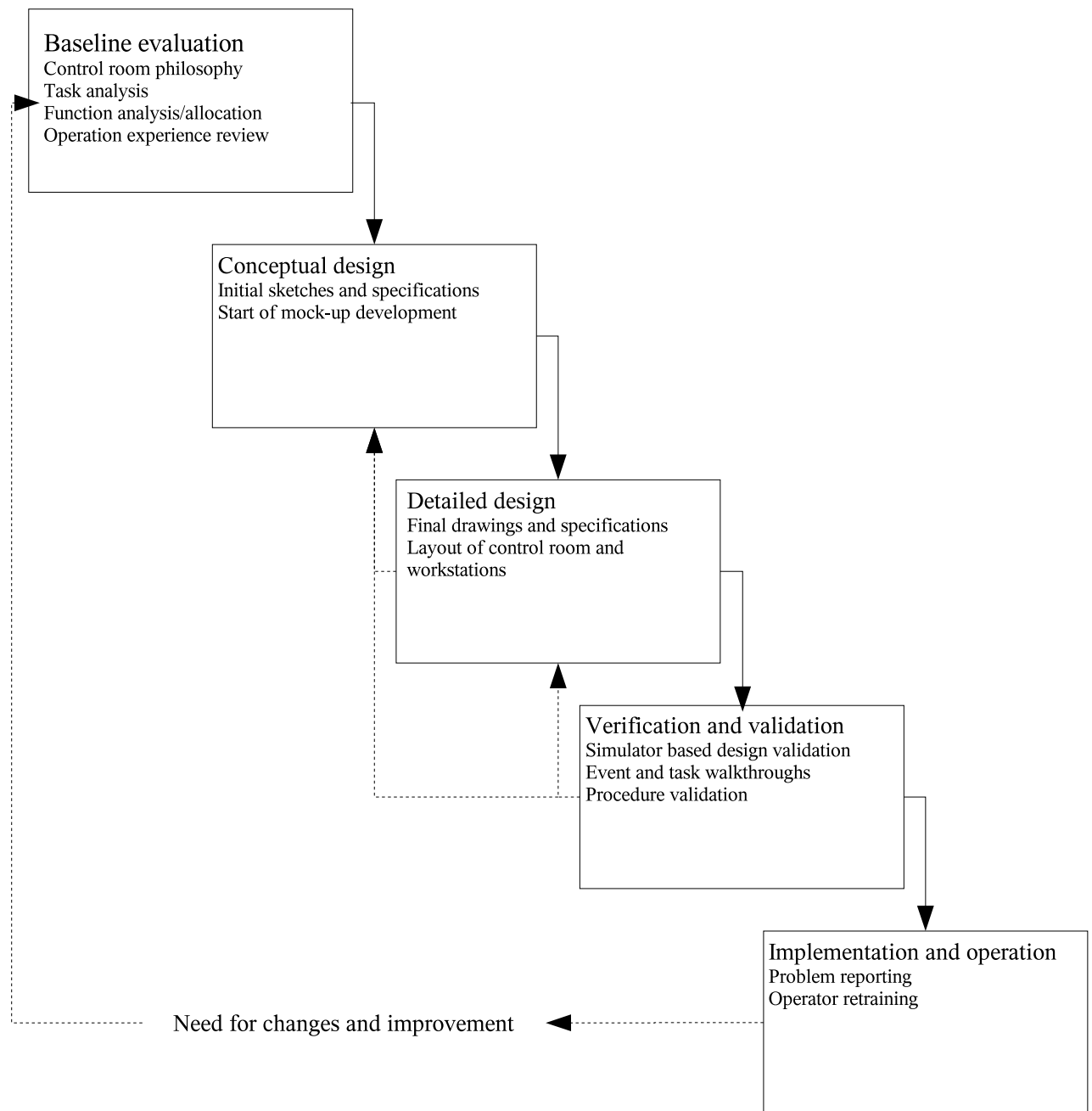


Figure 5. The connection between engineering psychology research in this thesis and a simplified human factors review model based on standards and guidelines (Drøivoldsmo and Louka, 2002; Geddes, 2002: IEC-964, 1989; ISO-11064-1, 2000; NUREG-0711, 2002).

There are several problems with this approach. In the design of a new plant, the man - machine system is a necessary function of what has been decided when choosing the process technology and plant systems to be implemented. However, technological, economical and political objectives are deciding the core system in a plant. Human operation of the chosen design starting with the baseline evaluation as shown in Figure 5 is normally applied at a much later stage. Even if choosing a type of basic technology where complex dynamic behaviour is likely to pose more problems for the operators, these problems are normally dealt with in a later stage of the design process (Papin, 2002).

Another problem is that there is a shortage of tools and knowledge supporting the designers in the early phases of control room design (Drøivoldsmo and Louka, 2002). The real experts on how to operate the plant are the operators who have their daily work in the control room. The operators are not trained to express their design solutions in computer assisted drawing (CAD) models and to judge the written proposals for how the new or refurbished control room is supposed to look. The human factors members of the design teams face the same problem. Without good tools for visualisation and the opportunity to perform practical testing on the proposed designs, the input from these groups in the design team is very limited (Ibid).

As far as human factors are concerned, it seems quite unrealistic to expect that designers will change their design options, in the case that human factor problems are detected in a later phase of (Instrumentation and Control or HMI) of the design. Without supporting the knowledge and techniques for dealing with this problem, the contribution to the design process from the human factors area is confined to a rather superficial domain.

Engineering psychology should be focused on the production of design-relevant knowledge. Tools and methods for gaining knowledge about and suggesting changes to correct potential problems must support the design process in a way that makes full utilisation of the operators and human factors staff expertise. At the same time, the methods must accomplish the criteria for scientific work. One way of doing this is by adopting some of the basic methodology from psychology in simulation.



## 4 SIMULATOR STUDIES

### 4.1 Simulation

An important contribution to what we today know as computerised simulation was laid at Kjeller and in Oslo, Norway, in the late fifties and early sixties. The development of SIMULA (SIMULATION LANGUAGE), a programming language developed by Kristen Nygaard and Ole Johan Dahl for the purpose of human understanding and computerised description of complex systems, laid the software basis for what we today know as modern object oriented programming. Through his work analysing soldier's tasks, Nygaard was confronted with discrepancies between data and real world observations. This led Nygaard to the conclusion that "data is a part of a social setting; they should also describe the reality of which they are produced, and the understanding of reality the users of the data possessed" (Haraldsen, 1999).

Since the mid-eighties, the use of simulations has grown rapidly. Simulations are today used in a number of areas, besides the engineering area, e.g., decision-making, training and planning. The use of simulation has become a common tool for problem-solving.

*The attraction of simulators is that they are easier and cheaper to use than real systems, e.g. simulation of car crashes. In many cases a simulation may also be the only way to investigate something either because the target phenomena may be out of reach, impossible to control, or simply because they do not exist yet (Hollnagel, in Cacciabue, 1998).*

For the process control area, simulation of safety critical events has become a central tool in both training and research areas where humans are a part of the system.

### 4.2 Simulators

Simulations are used for a wide variety of purposes: pilots flying aircrafts, military tasks such as maneuvering tanks and leading artillery fire, air traffic control, ship navigation and control, and in a number of process control settings for chemical and nuclear plants. Depending on the purpose of the simulation, simulators are built to resemble the task or process in more or less detail.

The use of simulators is today the main training and learning method for most of the areas listed here, and they will probably also be the main method in the future (EATMP Human Resources Team, 2000). However, several new areas of use have evolved. The area of verification and validation is growing in pace with more focus on the human operator of the systems. In the nuclear industry, typical simulator-based HMI validation experiments have been used when building new and refurbishing control rooms (Humbel, 2002; Roth and O'Hara, 1998; Braarud and Nystad, 2002).

Simulators for training and learning can be classified into 3 to 4 categories. Eurocontrol distinguishes between hi-fidelity simulators, simulators, part-task trainers and other training devices (EATMP Human Resources Team, 2000). The hi-fidelity simulators are full scope replicas of the real control rooms and control stations including all computer

programs and functionality found at a real control centre. In the process control area, a distinction is drawn between full scope and part-task simulation (Theureau, 2000).

The full-scale simulation refers to what is judged as a close to perfect copy of the real plant. For the nuclear industry, full-scale simulators are built to represent all functions in a control room and the simulators are used in the education and retraining of operators. When refurbishing the control room, the same changes are made in the simulator, and it is used for validation of the new systems and operator training. What the full-scale simulators often lack is an effective data collection system for use in engineering psychology tasks.

The part task simulation has, due to recent technological development, become more close to the full-scale simulation, but they do not cover all systems found in a control room. With respect to some features, they can be said to be even more realistic (see extending the part task simulator VR simulation below). Using Theureau's classification, the HAMMLAB is an example of a highly advanced part-task simulator. Some simulators in HAMMLAB cover most systems found in the full scale simulator, but the physical layout and means of control are different.

A micro world simulation is one example from the category of other training devices. The experience suggests that many fundamentally important systems are difficult to describe or control using traditional methods (Bargiela, 2000). These simulations today are not even close to a realistic interface, hence not directly relevant for this discussion, but the tradition will probably develop rapidly as a function of new technology. Utilisation of VR technology and modern immersion techniques can in the near future make interaction in fields like fire fighting and military operations highly realistic. The VR simulations are arguably the most spectacularly successful commercial applications of simulation, particularly in the domain of entertainment. The commercial gaming industry has been able to gather hundreds of thousands of people in the same virtual world with the widespread availability of the Internet.

### ***4.3 Extending the simulation***

To support the V&V process, a combination of full scale or part task simulation and features from the micro world simulations could be useful. The idea is simple. In addition to the existing interface, the simulator can be coupled to virtual or augmented worlds simulating the physical entities of the surroundings. These surroundings could be either inside or outside the control room.

#### **4.3.1 HAMMLAB as an example**

What really distinguishes HAMMLAB from a full scale simulator is primarily the philosophy and capability for development and testing of new data handling and presentation concepts. Physically, HAMMLAB is not a direct replica of any nuclear control room, but it is capable of representing a potential future control room design. For the nuclear units simulated in HAMMLAB the functionality of these designs allows for operation similar to the operation performed in the real plant. For some tasks, this simulation is even more advanced than the training simulators made specifically for the

actual plants. The software systems controlling the flow of data and interface design are flexible in such a way that most types of known process control representations can be implemented and used for practical tests and experiments. With the graphical user interface system Picasso-3 (Jokstad and Sundling, 2002), changes to the control interface can be changed and updated (off or on-line) for experimental simulations. Currently three different NPPs are simulated in the laboratory and the software infrastructure makes it possible to easily switch between the interfaces from one



*Figure 6. Picture from the HAMMLAB with the reactor operator control station to the left and the turbine operator control station to the right. A large screen is placed between the two control stations.*

simulator to another on any of the monitors shown in Figure 6 below.

### **4.3.2 Control room example**

As discussed above, the design verification process could gain from using VR mock-ups in verifying that the different parts of the control room are designed according to standards and guidelines. The last step in the human factors review models, before implementation and start-up of the plant, is the integrated validation. Undoubtedly, the resulting design could benefit from validation studies early in the design process. A next natural step of doing this is to combine VDUs and VR models in a semi-immersive control room environment such as systems used for e.g., air traffic control (ATC) and pilot training. In a flight simulator the cockpit is built as it is in the airplane, while the terrain is made up by a computerized representation. Adapting this technique to process control would involve a VR representation of the real control room as it appears on the plant, and VDU-based control stations replacing the control panels to be upgraded. An example of such an implementation is presented in Figure 7, where the conventional controls for the turbine side of the HAMBO (HAMmlab BOiling water reactor) simulator in HAMMLAB have been replaced by a VR-model in addition to the new VDU systems built to replace modernised control panels. This set-up allows for an extended simulation of the time, location and integration characteristics of the situation inside an advanced mock-up of the control room.



*Figure 7. Manipulated picture from HAMMLAB where a VR model is used to represent the conventional panels on the reactor side.*

Using the semi-immersive technique can allow for more detailed and hence more realistic operation in the control room earlier in the design process. Integrated system validation (O'Hara, Stubler, Higgins and Brown, 1997; O'Hara, 1999) could especially be started earlier. Upgrades of systems in the control room implies more automation which leads to changed procedures, working routines and a different distribution of functions allocated to humans and computerised systems. Combining the computerised interfaces with virtual mock-ups connected to a simulator will allow for testing of how computerisation of a system will influence the operation for both the operator directly affected by the upgrade, and the performance of the whole operating crew.

### **4.3.3 Field operator example**

Introducing wearable computers for the field operator raises many of the same questions as the change from conventional to advanced control room systems. In general, process plants are highly automated. Outdoor operators need to supervise, or to control the execution of specific tasks and need access to the control and support systems. During field operations, operators could benefit from computer support for monitoring and intervening the process. As discussed in Chapter 6, creating a test environment of the process plant, using the same principles as described above could be a way of providing access to a test bench.

The illustration in Figure 8 shows an operator working in a virtual representation of a process plant. Coupling both the VE and the wearable equipment to a simulator of the plant would open for testing of the wearable HMI as well as application content presentation.



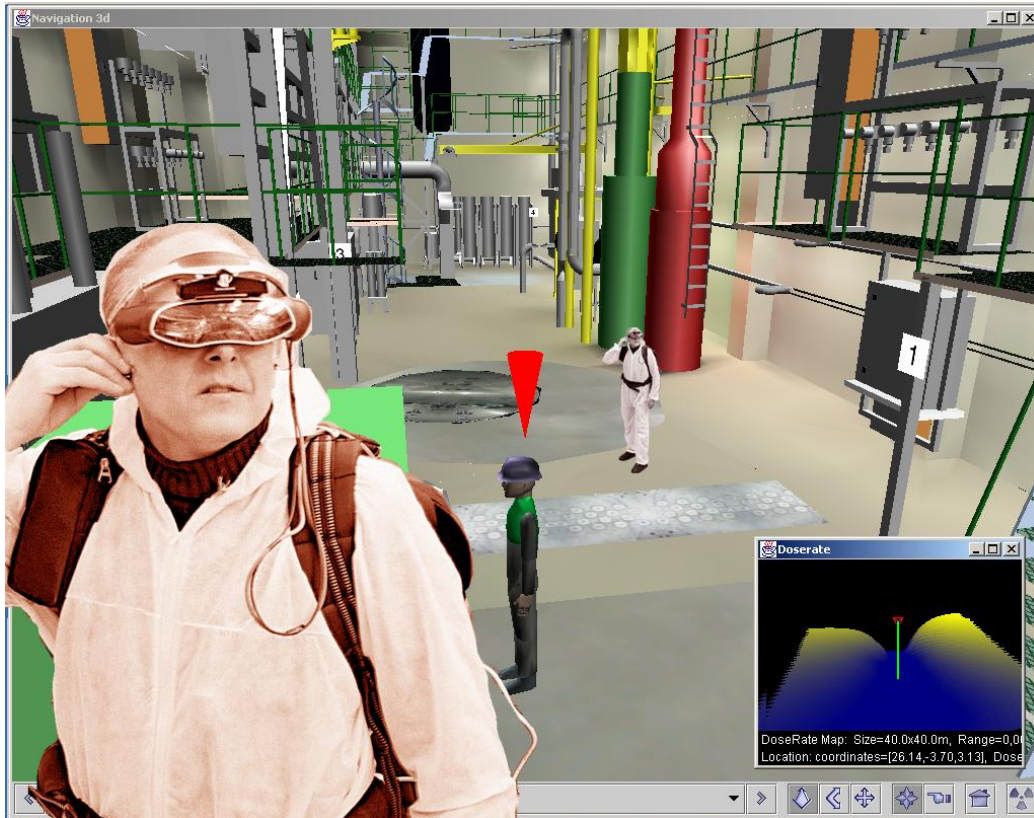


Figure 8. Field operator working in a VR model representing the process plant.

#### **4.4 Simulators as a way of gathering knowledge about human problem-solving**

For many years, cognitive psychology has been the main source of input to user interface design in complex safety-critical industries. The use of well-established principles for perception, memory and cognition has resulted in numerous volumes of guidance for how to design the operator's user interface, and an almost obligatory use of "psychology expertise" in various industrial projects. Nevertheless, the benefit from this expertise is not always clear, and the results from the use of the principles can often be questioned. One possible explanation of this could be the gap between the level of explanation needed for the design, and the level of explanation of the basic psychological theory. Pylyshyn (1991) argues that improvements in practical arts never have been proposed because of relation to validated scientific theory, and Carrol (1991) suggests that this situation is not unique for the relation between psychology and design. "It has been the general case that basic science provides uncertain and indirect support to applied work" (Ibid). The role of basic science in application is far from clear. Detailed articulation of the role basic psychology plays in HMI-design and verifying it in actual practice has proved difficult. The most sustained, focused and sophisticated attempts to develop explicit extensions of academic psychology for HMI have had no major impact on design practice. On the contrary, some of the best and most effective design has been done without any explicit use of psychology at all (Ibid). This could be summarised as: The need is in development of theory and principles to summarise regularities at a level appropriate for the design task.

Theureau (2000) describes the control room in a NPP as a living, social and cultural complexity. He claims that this system can not be broken up into simpler sub-systems and studied separately. In his study of navigation on a US Navy ship, Hutchins (1995) states:

*Cognitive science was born in a reaction against behaviourism. Behaviourism had made the claim that internal mental structure was either irrelevant or nonexistent - that the study of behaviour could be conducted entirely in an objective characterisation of behaviour itself. Cognitive science's reaction was not simply to argue that the internal mental world was important too; it took as its domain of study the internal mental environment largely separated from the external world. Interaction with the world was reduced to read and write operations conducted at either end of extensive processing activity. This fit the computer metaphor very well, but it made the organisation of the environment in which thinking took place seem largely irrelevant (p. 371).*

In his description of the cognitive approach, Neisser (1967) made this point simple and clear. Describing the classical procedures of experimental psychology as "brute force" (p. 305) and arguing for the view that circumstances constitute an important factor for higher mental processes, he pointed out an important gap between the laboratory research and the real world multiplicity of motivations and flexibility of response. The potential consequences of a gap between the human in an environment and the models of the mental inner world can not be afforded in industries depending on a high degree of safety. Even if the metaphors still are resident in the world of computers, the contextual dependence in the interconnection between human and machine and the task specific needs in the situations under study are taken more into account. Both in basic and in applied experimental psychology there has been a shift from the 'small and simple paradigm' to the analysis of complete tasks and skills (Sanders, 1991). The view of the Gestalt psychology, claiming that "the whole is more than the sum of the parts, and the properties of the whole can not be reduced to the properties of the separate parts" (Lübcke, 1983) has been re-discovered by the research communities in parallel with the technological development of simulators.

#### **4.4.1 Validity**

Validity is typically defined as the extent to which a measure assesses what it is supposed to measure. Normally three kinds of validity are considered: construct, content, and criterion validity. Construct validity concerns how well the operations being measured fit with the abstract, theoretical construct that the operations are meant to represent (Nunnally & Bernstein, 1994). This fit can be evaluated by means of the multitrait, multimethod technique (Campbell & Fiske, 1959), investigating to what extent the referent measure, i.e., the measure being validated, converges with or diverges from other measures (Cook & Campbell, 1979). If the referent measure correlates with theoretically related measures, then it exhibits convergent validity. Divergent validity, on the other hand, is evident when the referent measure does not correlate with theoretically unrelated measures. To ensure that the correlation between measures is not simply due to methodological similarities, the testing should be made using different data-collection methods (Nunnally & Bernstein, 1994). For example, to validate a measure of workload that is based on a questionnaire, this measure should be correlated with a workload measure using a somewhat different data-collection method,

e.g., psycho-physiological method. It is then reasonable to explain co-variance as being the result of conceptual overlap rather than methodological similarities.

When a measure is validated by means of a criterion, it is not what the measure is assessing that is of major concern, but its ability to predict some criterion (Kerlinger, 1986). The difference between convergent validity and criterion validity is that for the criterion-oriented validation, the referent measure does not have to be theoretically related to the criterion. It is difficult to perform this kind of validation within human-machine research, since many factors typically influence the criterion measure (Fracker, 1991). This problem is evident within human-machine research. Here one would like to identify the extent to which various factors contribute to the likelihood of human errors. However, isolating the contribution of individual factors is difficult because several factors typically act in parallel (Reason, 1990).

While criterion validity is performed by correlating the target measure with something outside the construct, e.g., to what extent workload predicts human error, a content validated measure is the criterion of performance (Nunnally & Bernstein, 1994). In human-machine research, measures of performance are commonly the criteria by which performance shaping factors are validated. To place confidence in such criterion measures, however, it must be ensured that the content is representative, i.e., that all significant parts of performance are encompassed (Skraaning, 1998). Determining the significant elements of the performance of a human-machine system will typically be left for subject matter experts to decide.

#### **4.4.2 Ecological validity and simulation**

The concept of ecological validity refers to development of theories that describe cognitive processes operating in realistic, everyday situations (Medin and Ross, 1992). Common criticisms of laboratory studies are that they are too artificial and not representative of the everyday situation in the real world. The other side of the issue is that ecologically valid situations are very likely (or even necessarily) experimentally invalid because one cannot tease apart the contributions of different correlated factors to performance (American National Standard, 1992, Medin and Ross, 1992). These claims are discussed further under the “experimental control” paragraph (4.4.3) of this document.

Sanders (1991) has argued that research interests have tended to shift from the analysis of traditionally limited and seemingly elementary behavioural phenomena to the study of cognitive and perceptual-motor skills with ecological validity as a major byword. Her concern is mainly the process control area and the utilisation of simulation, but there is hardly any contradiction between this area and the wider area of other behavioural sciences.

The effect of the computer revolution has so far only touched the surface of empirical behavioural studies. Utilising the full scope of opportunities provided by the computer will allow a new setting where real life tasks can be simulated in detail, not only the specific task itself, but also the natural surroundings of the task. In the industries where simulations are used, the concept of ideal simulation or full-scale simulation usually refers to a simulation of the physical properties of a production process or capabilities of a vehicle. An example could be a process plant where operator control monitoring and

problem-solving are natural tasks. Sitting in a large central control room the operators can control almost the whole plant. The traditional way of simulation has been to give the operators the appropriate feedback of plant behaviour when they perform various control operations. Experimental staff acts as the rest of the world outside the control room, communicating with the operators through direct speech and through audio communications. When operators in the control room ask the outside world for services or information, they get feedback based on the experience of the experimental staff acting in the scenery. More data power and advanced simulation will allow not only this kind of test situation, but it also involves the surroundings of the operator and the actual plant itself, i.e., building a virtual model of the plant simulated. This approach allows the whole plant to be simulated with highly realistic visual feedback, better time estimates for physical changes in the environment, over-heated areas not accessible in an accident, crowded areas making it difficult to solve necessary tasks, etc.

In addition to the representation of the surroundings, there are three important factors greatly influencing the success of achieving ecological validity in a simulator study. These three are the task (stimulus material or the problem), the subjects used in the study, and the data collection methods. All three factors are discussed in the subsequent chapters.

#### **4.4.3 Experimental control**

One way of describing the research methods in this area would be to look at the degree of experimental control. Experimental control is the ability to eliminate the effects of irrelevant variables by controlling them, leaving only the experimental variable free to change. In the situation of high degree of control you find the classical laboratory experiment, with the purpose of artificial separation of variables which are found together in natural settings, so that it can be determined whether one (or both) are important to an outcome. To satisfy the demands of the scientific method the following three characteristics are important: "The ability to change the independent variable, the ability to randomise other variables, and the ability to eliminate (or support) alternative hypotheses through systematic variation in a series of experiments" (Lewin, 1987).

Changing the independent variable is not the problem. The challenge is to control the other variables in a systematic way that allows testing of the hypothesis. In a realistic complex task, strict control or randomisation of all variables influencing the problem under study is practically impossible. Even trying to keep full detailed control would be a contradiction in terms. It would prevent the subject from free choice of problem-solving strategies and then influence the realism.

On the other hand, what can be controlled is the outer experimental surrounding of the simulation shown as the top level in Figure 9. The experimental surroundings are made up of the initial situation, the way the experimental apparatus is applied, and the way the experimental staff controls the speed and accuracy of the simulation. What can not be controlled in the initial situation is the subject's knowledge and training and unpredictable interactions between these factors and the subjects' apprehension of the experimental situation. The surrounding level lies outside the subject's interaction with the simulator, the simulated world, and of course outside the subject's internal cognitive

processes. Applied in a non-intrusive way, these surroundings could be controlled through good experimental procedures.

On the next level of Figure 9, we find the simulated world. This level consists of the actual simulation and the tools available for solving the problem, i.e., different support systems for helping the operator perform various operations. For the control room operators the simulation includes both the HMI to the simulator and the VE comprised by VEs connected to the simulator. What can not be simulated has to be acted by the experimental staff. In a well-simulated environment, this level is the easiest level to control. However, using a simulator to study problem-solving in a realistic manner, demands that the simulator is operating close up to the real world simulation. The operator should have the same feeling of time pressure as in the real world task (Brehmer, 1995). Traditional laboratory tasks are often static tasks, where the subject has unlimited time to make decisions. Advanced simulator studies are not static, but can be prone to considerable delays in output and feedback to the operator. This feedback delay is of course not important as long as it is representative for or similar to the delay experienced by operators in the real world. In this case the feedback delay is a necessary feature of the system, but experimental simulators are often burdened by extra logging and data collection systems, and the problem has so far been to produce fast enough simulators to take the simulation up to the speed of the real world system. For the use of VR models, this is an even bigger problem (as described in chapter 6.1).

The inner circle in the figure represents the topic of study: the operators acting in the artificial world, solving problems with their knowledge, and the process interface available. As described by Funke (1995) "the experimenter merely sets the stage for a subject who then follows an idiosyncratic path". In an experiment making use of simulation, as opposed to the controlled laboratory experiment, the operator gets no detailed instruction about how to solve his task. The operator knows this beforehand from several years of training and practice in the real world.

The experimental control in this situation does not rely on detailed instructions about how to perform the experimental task. One relies on an assumption of the subject's approximately constant behaviour in realistic surroundings. Still, a challenging problem is keeping control of the scenarios. In a dynamic world, with a high number of interacting parts of a complex system, where every step in a series of interventions can generate a special situation, it is very optimistic to believe that scenarios can be designed in a way that allows identical runs in the simulation. At best, the experimentalist can aim at designing situations where the operators are solving the same problems in approximately the same sequence of significant events. There are several ways of approaching this problem.

The most common way is to define events that are inserted into the scenario by the experiment leader. The events should be identical for all subjects and appear at approximately the same time in the scenario. The problem with this time-driven approach is ensuring that the working context of the events are comparable between crews (Jorna, 2000). The working context depends on the speed and accuracy of the different crews and using an event-driven approach focuses more on similar context. One obvious problem with the event-driven approach is the number of possible actions leading to an almost idiosyncratic path for each crew in every scenario.

A requirement for success in this approach is that the subjects are from the population of people that are normally exposed to the problem type under study, e.g., studying problems in complex process plant surroundings using university students weakens the experimental control. Using untrained subjects can only be used for the purpose of investigation of how untrained subjects react to certain stimuli. In other words, the subjects have to be representative for the population where the situation is occurring and act as they would do in their own well-known plant (Follesø, Drøivoldsmo, Kaarstad, Collier and Kirwan, 1995).

Being used to the situation of simulation is also necessary. NPP operators normally accomplish retraining in a simulator once a year. This training is performed in full scale simulators and is a necessary and natural part of their working situation. Other domains like ATC and aviation have similar routine simulator training. The training performed in these training simulators is their way of learning to control the real process, and this control depends on the operator's ability to perceive the simulated situation in the same way as the real situation in the plant. The main point of the experimental control is that the task of controlling the plant overrules all other effects of the surroundings. As long as the situation represents the natural situation in the control room, instructions to the subjects are mostly superfluous. Regardless of the instruction, the subject will act as s(he) would in the real situation.

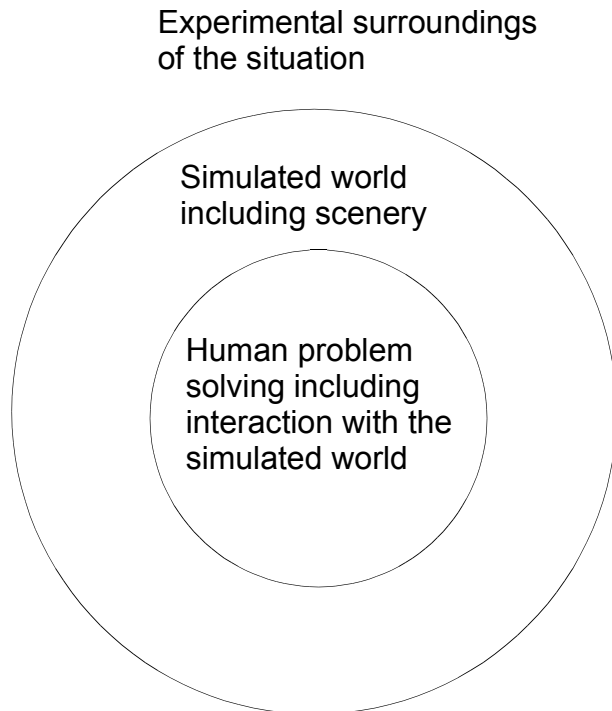


Figure 9. The world surrounding the subject in a simulation is described as the immediate surroundings consisting of the simulator and the more peripheral experimental control mechanisms.

The operator's insight and willingness to act in the simulation can easily be disturbed. Experience from the HEAP (Human Error Analysis Project) described in Drøivoldsmo (2000) indicates that even small simulator discrepancies from the real world situation influence the expert operator's performance. Experts usually know their plant's behaviour down to the details and unexpected plant behaviour could lead the operators to withdraw from the situation. Instead of solving the problem presented for them, the expert operators could start focusing on the flaw in the simulation.

Making a distinction between the external and the internal factors constituting the experimental situation (as in Figure 9), one could distinguish between the classical detailed control level and a higher level of control. The higher level of control is based on the requirements stated above and is believed to be a function of a high degree of task dependence in complex systems acting as in the natural world. As illustrated in Figure 10, this could make it possible to get more realism in situations of high experimental control.

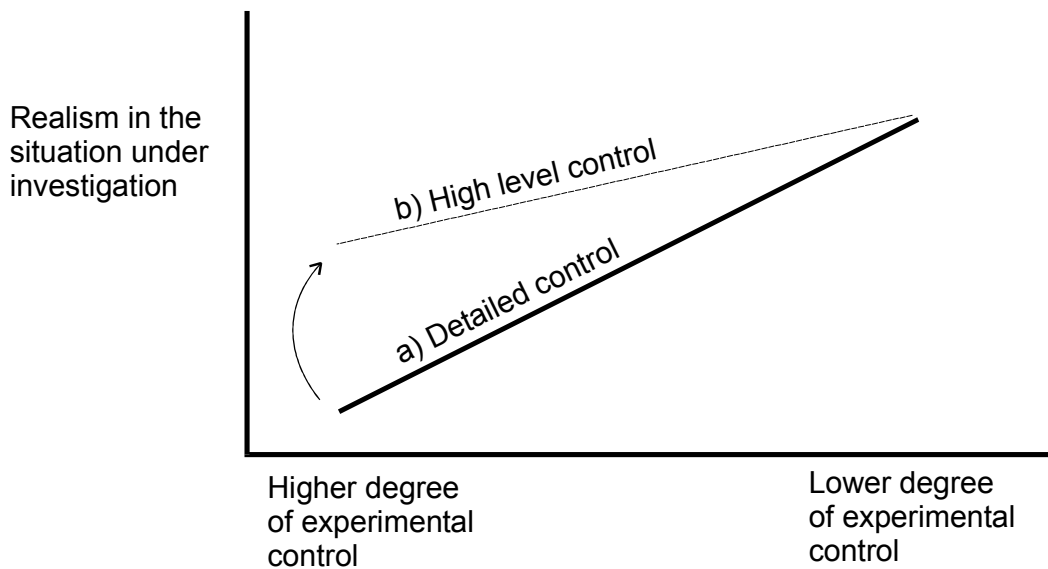


Figure 10. The solid line in the figure shows the traditional view that the degree of realism is low in the case of high experimental control and high in the situation with low degree of detailed experimental control. The stapled line shows a proposed situation with high level control of the experiment giving better realism.

Depending on how controlled the experiments are, or how “real” the situation in field studies seems to be, data collection and choice of methods are de facto standardised. The reasons for this rigid choice of methodology are several. Research communities often work in the same traditions and research areas, and people are recruited into a tradition; they are often loyal to the methodology the tradition offers etc. Following this, the development of new methodology is strongly influenced by and regulated by the leading researchers in their fields. It should be safe to state that this conservatism does not directly accelerate development of measurement methodology.

#### 4.4.4 Close to everything can be simulated, but not yet - an example

Reither (1981; cited in Funke, 1995) validated a third world simulation by comparing the performance of novice users, represented by post graduate students, and experts, represented by technical advisers that had more than ten years of practical experience from advisory work in third-world countries. Both groups worked in teams of three, and the sad outcome of the study was that for both groups the number of inhabitants in the simulated country decreased due to starvation. However, the results from the study concluded that experts showed a broader range of actions and willingness to make decisions from the start than the novice group. Intuitively most people would say that the same performance results of expert and novice behaviour probably do not represent the results we find in the real world. The results could have several explanations, but one of the following is the most probable.

1. The simulation model was not good enough, i.e., the results from manipulation of parameters in the simulation did not represent the results one would obtain manipulating the same variables in a real world situation. This could be caused by lack of a valid model due to the tremendously difficult task of simulating something as complex as a society.
2. Manipulation of the simulation was different from the situation of advising people out in the field. Lack of consistency with the real world interface would not allow the expert group to manipulate the variables they wanted to manipulate, led the operators to make mistakes and reduced their awareness of the situation the inhabitants were in.
3. The model and the control mechanisms were good enough, but results from technical advisory work in third world countries do not necessarily lead to better conditions and fewer deaths of the inhabitants.

Following the criteria outlined in the preceding paragraphs, these results could have been different. In a simulation of a NPP, e.g., like the one in HAMMLAB the representation of the interacting physical processes and control systems are close to accurate. Using professional operators familiar with the situation of simulation, manipulation of the different control mechanism is usually no problem. In a study comparing professional experienced operators from the Loviisa power plant with operators unfamiliar with that plant, Follesø et al. (1995) found the professional operators to be both quantitatively and qualitatively different. The operators with experience from the plant simulated outperformed the unfamiliar operators. Finally, the knowledge of the mechanisms controlling the real world behaviour needs to be known before expecting the simulation to produce valid and reliable data.

To sum up, it should be documented that a simulator is capable of simulating the world in a way that makes us trust the results. Results should not be expected after short time simulations, unless the simulator is close to the subject's real world background.



## 5 HUMAN PERFORMANCE ASSESSMENT IN SIMULATION

Research methods in the area of psychological engineering and human factors in process control struggle with a measurement problem. As described by Lewin (1987) this is not a problem limited to this particular area.

Because methods of measurement are so important to science, scientists have a strong tendency to study only what can be measured with the methods already in existence, so this problem is more or less universal, but experimental studies of behaviour seem to suffer more than other sciences.

Good general performance indicators for operation of a complex system are hard to find. There is often more than one correct solution to a problem, and even subject matter expertise with long experience can find it problematic to rate and to differentiate the quality of good and poor performances. Actions that seem correct and plausible at a point of time can be disastrous in the next moment or hiding a dangerous error two hours ahead. Traditionally most psychological experiments have been performed with single subjects using speed and accuracy as the dependent performance variables. Controlling the presentation of the stimulus, the experimenter can register the time from stimulus onset until the task has been finished by the subject. The equivalent to these measures in complex simulations is identification of critical subtasks and registration of time the same way as in the traditional controlled experiment. However, this can be burdened with problems as long as it is not certain which event or combination of events in the simulation triggered the operator to perform a certain action. The time needed for simulation of a problem could also make the establishment of stimulus response (S-R) relations difficult. In slow industrial processes, the time span of a realistic simulation could be several hours, in extreme cases several days. Such long time intervals allow the operators to test and discuss different strategies for solving the problem and make it difficult to interpret number of seconds or minutes as indicators of performance. High amounts of time spent in such situations could be interpreted as slow operation as well as safe and thoughtful decision-making.

When the criteria for performance are less clear, the classical dependent variables measure on a low level, i.e., time and correctness are far from sufficient for rating performance. Systematic experiments in an ecologically valid setting as full-scale simulations, often require measurement methods adapted to the situation.

### ***5.1 Complex problem-solving***

Characterising performance means systematic descriptions, categorisation, and discrimination between aspects of performance (Moray and Huey, 1988 p. 73). To be able to identify underlying causes of quality of performance, performance should be broken down into parameters related to a theoretical model allowing comparison of performance measurement over a variety of contexts.

Dynamic problem-solving in laboratory simulations has been described with three characteristics by Brehmer (1995). (1) They require a series of independent decisions, (2) the state of the task changes, both autonomously and as a consequence of the decision maker's actions, and (3) the decisions have to be made in real time. Complex problem-solving in a NPP simulator would satisfy all these characteristics. An example of such a task could be a control room operator getting an alarm from the plant. He starts looking for the cause of the problem and detects reduced level in a tank. From this information, he concludes that he has problems with a pump and starts to plan an action to correct the problem. Meanwhile an automatic protection system starts preparing the plant for automatic shutdown. The situation described is dynamic in the sense that any action the operator decides to do to correct the problem will influence the controlled nuclear process in different ways, leaving the operator with a set of new decisions to be made. The process will autonomously develop in one or the other direction, depending on the operator's choice of actions and the point of time different actions are performed.

Funke (1995) suggested a taxonomy for variables that affect complex problem-solving consisting of three factors. (1) Person factors, comprising the competence that a subject introduces into the problem-solving situation. (2) Situation factors representing instructions to the subject during a scenario. (3) System factors consisting of both the formal and content related attributes of a system. In his description of system factors, Funke uses the concept "Eigendynamic" which is a change of state in the system simulated at a given time  $t$  due to the values of some variables at an earlier point of time  $t-1$ . This change takes place independently of any input by the operator. The influence of "Eigendynamic" in a complex process plant impacts how human performance should be measured. Eigendynamic in a process plant could be interpreted as the response of all automatic systems and fluctuations due to physical processes. A serious implication for performance measurement is that actions performed by the operator at a given time result in a chain of events directing the future development of the process in a way that may not be visible because of feedback delay. In this "dead time" (Brehmer, 1995) the operator can perform a number of critical operations leading to a positive or negative effect on the future. Operational decisions/behaviour dangerous for the plant can be corrected before any serious damage has occurred, and they may not have impact on the total outcome of the simulation in sense of reduced power production or damage to the plant.

Examples of other situations could be scenarios where the operator performs a series of operations with negative effect on the plant, but the plant automatic protection systems correct the problems with or without the knowledge of the operator; or, where the operator makes the wrong decision, but never manages to realise the plan. The opposite situation is of course equally relevant. "Even when plant performance measures are maintained within acceptable ranges, shortcomings in the design may result in unnecessary demands being placed upon the operators" (O'Hara et al. 1997). For example, the situation where the operator found a good solution to the problem, but an inappropriate design obstructs the effectuation of a good plan.

For the purpose of design, human performance measurement should be able to account for all these situations described above. Therefore, a model taking account for a number of relevant parameters is needed.

In what he calls the psychological model, Skraaning (2002) describes human-machine interaction as a process where the operator's problem-solving impacts the system within the following causal chain:

System input → operator cognition → operator behaviour → system outcome

This psychological model separates the operator from the technological system and specifies the causal parameters behind human intervention. Skraaning (2002) further points out that:

*the causal process underlying the psychological model is a static and probably inappropriate description of human-machine interaction. In addition, the narrow focus of the model overestimates the importance of the human operator for the control of advanced technological systems. Being aware of these weaknesses, the psychological model will still be considered a useful theoretical framework for human performance measurement in complex operating environments*

Extending this model with the relationship to the technical system and a regulation mechanism, presents a more complete model in the sense of better representing the real world situations, but it also complicates the interpretation for the purpose of human performance measurement (Figure 11). However, to handle the situations described

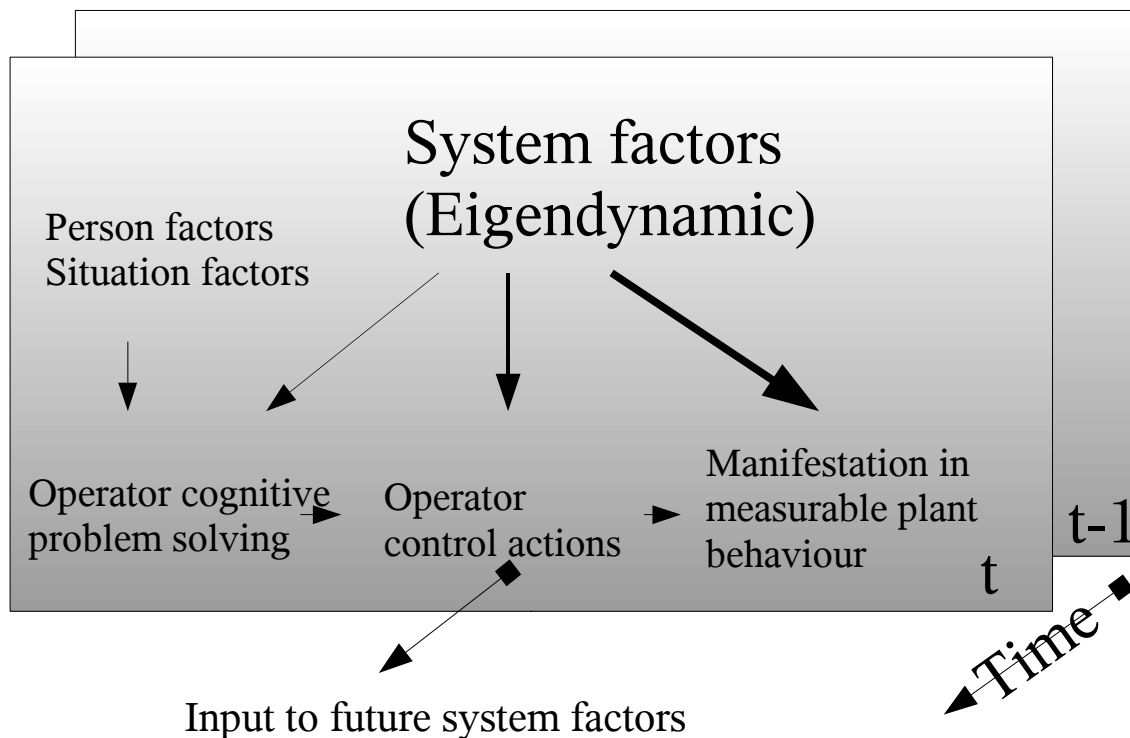


Figure 11. A model useful for describing human performance. Brehmers' characteristics of complex problem-solving and Funkes' taxonomy for variables affecting human problem-solving applied on the three different measurement levels described by Skraaning (2002).

above, the more complete representation may seem necessary. The representation of the time dimension is necessary for the capture of design relevant perceptions and interpretations of the operator not necessarily manifest in observational behaviour or

technical system responses on the cognitive problem-solving level described in the model. The system Eigendynamic is needed to account for the changes in the process not directly controlled by the operator, and the situation factors are the influencing features of the given situation like, e.g., instructed use of operation procedures.

One implication directly deduced from this discussion is that design relevant performance depends on measurement techniques able to produce data for short time windows during the simulation. As described below, process tracing techniques are relevant indicators for this purpose.

## ***5.2 Selecting indicators of human performance***

In addition to the validity issues discussed above, the following factors should be evaluated when deciding upon indicators for measuring human performance (Andresen and Drøivoldsmo, 2000; Gravelle and Shopper, 1993; O'Hara et al., 1997).

### **Level of Detail and Sensitivity**

The method must provide data that have the appropriate level of detail for meaningful analysis (American Institute of Aeronautics and Astronautics, 1992; Gravelle & Shopper, 1993). Inappropriateness can be the result of too little or too much detail. When data-collection methods provide too much information, refinement methods can be applied, aggregating or filtering the data into the needed level of detail. For measurement methods, the question of level of detail is closely connected to that of sensitivity. Sensitivity means that a measure has the ability to reflect all fluctuations of the construct being measured. In performing experiments, one should select measures that are known to be sensitive to the experimental manipulation. Insensitivity appears as ceiling or floor effects (a large quantity of the participants get the highest or lowest score), limited range (only a subset of the possible scores are obtained), or other unwanted distributional characteristics (Skraaning, 1998).

## **Reliability**

A reliable method yields the same type and amount of data when applied under similar conditions using the same participants. Reliability is indicated by a coefficient ranging from 0 to 1.0. A reliability of 1.0 means that the measure is perfectly reliable, whereas a ratio of 0 means that the data being compared are uncorrelated. Two ways of testing the reliability of a measure, are to see whether (1) the measure is stable over time, or (2) the items within the measure are homogeneous (Kerlinger, 1986). The former is often referred to as test-retest reliability and is obtained by correlating one measure across different occasions, using the same participants under similar conditions. The latter is termed inter-item reliability and is acquired by splitting the items into two subsets and correlating all possible subsets to one another (Cronbach, 1990). When the data are collected by means of observer ratings, the rating may vary somewhat from observer to observer due to, e.g., slightly different interpretations of the criteria by which the rating should be performed. An inter-scoring reliability coefficient can be calculated to find out to what extent the observers agree. Subjective judgment and reliability are closely connected to the criterion of objectivity described next.

## **Objectivity**

Objectivity refers to the extent to which a method provides data that are unbiased by subjective judgment. Objectivity of psychological measurements is primarily a question of how the data are collected. For instance, ratings and expert judgment are methods that rely quite heavily on subjective judgment. However, the objectivity of these methods can be improved by developing clear procedures and performing training sessions (Cronbach, 1990). When the verbalisations of the operators are used as data, it is possible to distinguish between two types of subjectivity: self-report and subjective judgment (Fracker, 1991). Self-reporting yields raw-data, in the sense that the participants only report what they are thinking ("thinking aloud"). Subjective judgment, on the other hand, implies that participants operate as measurement instruments, assessing their own thought content (Meister, 1985). Ericsson and Simon (1980) identified a number of ways to collect verbal data, and the type of biases to which the various methods are subjected.

## **Intrusiveness**

A data-collection method is intrusive if it influences the operator's performance. The method may affect performance by providing cues, interrupting the sequence of actions, or increasing the load on the operators. Concerning cues, questionnaires may indirectly reveal the purpose of the study, guiding the performance in a specific direction. Posing questions during freezes of the scenario may interrupt the natural flow of the problem-solving process, although Endsley (1995) does not agree with this view. Asking the operators to report on their problem-solving may have different effects on performance depending on whether the reporting is done after, or concurrently, with the experiment (Ericsson & Simon, 1980). If the information that the operators are to report is not readily available, concurrent verbalisation may increase the load on the operators and slow down performance (Ibid). Also eye-movement tracking (EMT) and psycho-

physiological methods, might under some conditions be intrusive, on the operators; e.g., the operators may become more easily stressed if the equipment is uncomfortable to wear.

## **Utility**

The utility of a method is assessed by comparing the value of the data obtained with the resources needed to apply the method. If the method is costly to use or requires special training or education beyond the researcher's main field, then the value of the data must be comparatively high. For example, when deciding whether to use technically advanced methods during an experiment, e.g., eye movement tracking, the method's utility should be considered. Eye movement tracking equipment is quite expensive, especially for the case testing wireless applications, and some training is needed to get reliable data. However, since the method yields unique data that might not be obtainable by other means, the extra costs and efforts might be worth it. Cost-benefit assessments are also required when deciding whether to develop a data-refinement tool. In the short run, the best solution often is to aggregate or rearrange raw-data manually. However, if the procedure is likely to be used in several experiments, it might be better to develop some computerised tool. For example, when performing experiments in HAMMLAB, the operator's interaction with the interface is being logged automatically. Since the logged data are of interest to most experiments, tools that assist the researcher in working with the logs significantly reduce the time spent on data handling.

### **5.2.1 The need for process expertise**

An important factor for successful achievement of ecological validity and experimental control (as argued above), is the development of simulation scenarios. The simulation scenarios are the way of testing the practical or empirical hypothesis. The scenarios are relevant operating situations that can be designed in order to test the hypothesis, such as the hypothesis of performance improvement due to an experimental manipulation of a given system. This process of scenario construction is highly critical for the outcome of the study. As shown by Federico (1995) and several studies of expert behaviour in chess, experts tend to be more context dependent than novices are. This fits well with the idea of using the known context of a familiar situation as basis for the study, but it introduces a new problem. The effect of the experimental manipulation could be overruled by the effects of the task.

Using realistic situations requires analysis close to the process simulated. In simulation of complex industrial systems, only on rare occasions can one person fill the role of both the experimentalist and the subject matter expert (SME). Implicit in the concept expertise lies a high degree of specialisation in a narrow field. People with a detailed knowledge of a process plant are, with almost no exceptions, educated and trained in engineering disciplines with focus on how to run and maintain the plant. As illustrated by Chalmers (1999) "The problem is not to find competent subject matter experts it is merely to train them in scientific thinking." To be able to take advantage of the SME, the training should not only be for the latter. To be able to communicate, the experimentalist also has to take his turn in understanding the process.

One important lesson learned in from the simulator studies in HAMMLAB is the need for a close co-operation with SMEs throughout the whole scoring and analysis phase of an experiment (Drøivoldsmo, 2000). There is no way the complex process as the one found in a NPP can be sufficiently learned and understood by research staff without hands on experience. Scoring and analysis where process knowledge comes into consideration should always be guided by subject matter expertise.

### **5.3 Data collection methods and measures overview**

Broadly stated, method refers to doing something in a systematic way. Typical methods used for collecting data in human-machine experiments are questionnaires, ratings, interviews, audio and video recording of the operator's behaviour, logging of the operator's interaction with the interface, and logging of the process fluctuations. Data-collection methods are simply the means by which activities of the operators are registered and transformed into a format that enables the researcher to perform systematic analyses. The collection methods, as such, do not carry with them instructions on how the data are to be interpreted. If the researcher would like the collected data to reflect quantities of specific psychological constructs, then the collection method must be part of a measurement method. (From Andresen and Drøivoldsmo, 2000.)

Measurement methods specify what type of data collection method to use, and rules on how a psychological construct maps onto the collected data. The collected data should then be conceived of as measures. Typical constructs being measured within human-machine research are, to name a few, workload, vigilance, stress and situation awareness. Each construct may be measured by means of various collection methods, e.g., situation awareness can be assessed by using performance data, self-ratings or responses to factual questions (Fracker, 1991).

If, on the other hand, the data collection is performed without any criteria on how to quantify particular constructs, e.g., automatic logging of the operator's interaction with the interface, then the resulting data are better conceived of as raw-data. This definition of raw-data is in line with what Coombs (1964) refers to as recorded observations, and not as data which have a restricted meaning and are ready for analysis. There are at least two important reasons to collect raw-data. Firstly, for some measures it is not feasible to perform the measurement on-line. This might be the case when the registered data go through several steps of analysis before the measure is obtained, or if diverse data sources need to be considered simultaneously. Secondly, raw-data are less constrained by the researcher's initial hypothesis than measures. That is, while raw-data should provide an accurate, unconstrained picture of what has taken place, measures are selected so as make the test of the researcher's hypothesis as powerful as possible. This means that raw-data can be a very useful supplement to measures when the study is of an explorative nature. Another situation where raw data is valuable is the case of quasi-experimentation as discussed in Chapter 2.2 of this document. In this case, a total re-interpretation of the whole data set can be done with the purpose of investigation of a new hypothesis.

However, since the amount and the format of the raw-data make them difficult to analyse, the data often need to be refined in some way. This is then the purpose of the data-refinement methods: to aggregate, rearrange and reformat raw-data. Just as for the data-collection methods, refinement methods do not produce psychological measures. They are simply tools for processing the raw-data. For example, when performing some sort of classification based on the operator's verbalisations, it can be easier to make the classification after the verbal protocols have been transcribed into a data sheet, rather than doing it while listening to a tape-recorder. The procedure used for transcribing the verbal protocols is a refinement method, since the aim of the transformation is simply to put the data into a different format. If refinement methods are used to produce measures, which they very well may (Andresen and Drøivoldsmo, 2000 p. 14), then they should be thought of as being part of a measurement method. The measurement method would then comprise at least one data collection method providing the raw-data, and one or more data-refinement methods yielding the measure.

### **5.3.1 Measuring operator cognitive problem-solving**

One way of data collection is to freeze the simulation, and collect subjective data from the subjects by use of interview or questionnaires. The advantage of this method is that the reports from the subjects can be collected repeatedly through experiment and different phases of, i.e., problem-solving can be evaluated together with the computer logs showing the status of the simulation at the times where data was collected. Unfortunately, people are not good at reporting detailed information about past mental events or even recent ones. "Recall is stilted by the amount of time and intervening events that occur between the activities of interest and the administration of the questionnaire" (Nisbett and Wilson, 1977). Retrospective verbal protocols (reporting events that occurred at an earlier point of time) are influenced by the biases caused by the subjects rationalising their behaviour. This procedure for data collection could also influence the ecological validity of a study. Even if repeated interruptions in a simulation scenario not necessarily have an effect on the subject's problem-solving (Endsley, 1995, 2000; Kaarstad, Follesø, Collier, Hauland and Kirwan, 1995), the contents of questionnaires and interviews traditionally applied in this kind of studies can have other effects difficult to foresee. Ideally, data should be gathered without taking the subject's attention away from the natural problem-solving situation, but the advantages and flexibility of the freeze techniques can not be overlooked.

Another way of collecting data is the concurrent, or think-aloud verbal protocol (CVP) technique. CVP tracing is a technique that was rediscovered in the late seventies after being banished for sixty years. CVP is verbalising information at the same time as the subject is attending to it (Ericsson and Simon, 1993). One drawback of this technique is that it is extremely time consuming (time for analysis is 20 to 60 times the raw tape), and burdened by a lot of prejudice. Another drawback is the problem with missing arguments when the situation demands many cognitive resources from the person verbalising. Studies of problem-solving in the HAMMLAB using simulation of a NPPs have shown CVP to "dry out" in the most interesting parts of the simulated scenarios, the part where problem-solving is most likely taking place (Braarud, Drøivoldsmo and Hollnagel, 1997; Hauland, 1996; Kaarstad et al., 1995).



Subjective data reporting, i.e., different workload measures, task complexity, and performance are biased by the operators for the reasons stated above. Different measures can seem to tap into one or more common variables and effects are difficult to isolate. Combining different measurements, like objective and subjective measure of the same phenomenon has shown promising results (Fracker, 1991). What is really needed is measurement based on effective scoring and aggregation of continuous data of high resolution. This way the critical time windows for S-R measurement can be determined post-hoc, and the consequences of the actions can be studied in relation to the actual content of the operator behaviour during the simulation.

One promising technique in this area is the use of EMT. During the last 5-10 years, technology for using the point of gaze has become more applicable. EMT undoubtedly has big potential, but the complexity of the technology and cost of data analysis has until now prevented spread of the technique as a tool in complex experiments. The strength of this technology is that fast and accurate information about what a person is looking at, for how long the object is looked at, and what the person looks at next, can be recorded with high degree of accuracy. A technique for use in process control studies is described in Drøivoldsmo et al. (1998). The major obstacle to overcome using this technology has been how to score and interpret data, but examples of successful application can be found. For example, in a full scale simulator ATC-study, Hauland (2002) successfully used visual attention strategies as an indicator of cognitive performance.

Recent development of the technology has allowed the combination of VR/AR environments and eye movements tracking. Use of head mounted displays with integrated EMT equipment (Shimoda, Ozawa and Yoshikawa, 2002) has recently demonstrated that data analyzable in a psychological tradition can be gathered in complex situations using this type of advanced equipment. Considering the fixed relationship between the head and the display, EMT in combination with HMDs could be less complex compared with, e.g., head-mounted EMT in a traditional NPP simulator.

### **5.3.2 Questionnaire based data collection using the Internet**

Computer-assisted psychological assessment has been widespread since the introduction of the Personal Computer (PC) in the 1970s (Cohen and Swerdlik, 2001). More recently, the Internet has been used for data collection in surveys and experiments. In the extended simulator experiment, as described in this document, the computerised test presented on the traditional PC-monitor does not fit with the needs of data collection. The computerised tools have to support a wide range of technical devices and different operating platforms for wearable equipment, PDAs, simulator interfaces etc.

One way of solving the problem is to adapt to the Internet standards and make the data collection system independent of the test-device. WWW-browsers are made available for most portable and hand-held devices as well as the traditional desktop and mainframe systems. There are a number of advantages to using computerised Internet based data collection. The advantages do not only apply to the speed and reliability of data collection, but just as much to the whole process of designing, training, informing and post experimental follow up of the test subjects. The full discussion of use of

Internet data collection and a system for performing such collection is discussed in Drøivoldsmo and Johnsen (2002). Table 3 gives an overview of the advantages of Internet versus paper and pencil. As shown in the table, the need for computerised support is different for the different project phases.

*Table 3. An overview of the discussion of advantages of Internet based systems for the project of running a simulator study with use of subjects in a geographical location different from the simulator laboratory.*

	Task	Traditional	Internet based
Pre-experimental work	Administrative and demographical data	Slow. Questionnaires can be left at home or stuck in mail	
	Training and introduction courses	Difficult to control	Allows time recording and monitoring of activity
	Last minute changes	Depending of project type	Can be done up to presentation of the first test subject
Experimental work	Control with missing data	Manual	Automatic
Post experimental work	Debriefing questionnaires	Slow	
	Debriefing tests	Difficult to control	Control possible with activity monitoring
	Data preparations	Manual	Automatic
	Raw data backup	Paper copies stored in different buildings	Save to different disks or backup medium located in different buildings

In the data gathering for the experiment described in Drøivoldsmo et al. (2002) the system described was used for collecting data on site inside the containment of the Halden Reactor. Using a HMD and a speech recognition system, questionnaire data was collected during the experimental test of a wearable system for AR radiation visualisation. The results showed the importance of data collection on site without time delay and interfering actions between the test and the data collection. Results show that the immediate responses given by operators inside the containment, gave higher knowledge scores compared to the test taken immediately after they were checked out of the plant. The immediate test was also superior to the delayed condition in regard to the time spent answering the questionnaires for both the knowledge scores and operators' subjective assessments of their own performance. The higher knowledge scores in the immediate condition were explained with memory decay and the operators' access to

environmental cues. In other words, removing the environmental cues in the test situation could jeopardise experiments claiming a high degree of ecological validity (Drøivoldsmo and Johnsen, 2002).

### **5.3.3 Operator control actions**

Performance scoring of operator control actions indicates to what extent the operator has succeeded in attaining goals, sub-goals and tasks necessary for successfully solving the problems presented in a scenario. The technique requires a pre-defined path for good operational practice, and the opportunity to do modifications to this path, based on SME evaluations of alternative solutions. Several objections could be raised against this type of performance measures. The most serious is the subjective bias by the process expert. The process expert will, at any point of time, base his scoring on his own epistemology. His training and problem-solving abilities will decide if he is able to perceive and interpret the full complexity of the situation. One example of a problematic situation will be the case where operators possess knowledge about alternative solutions, not known to the expert. In this case, the expert could be led to think that the operator has erred in omission or commission, when the operator has actually solved the problem through another strategy or route compared to what pre-experimentally had been known as possible solutions. Skraaning (1998) developed a computerised scoring system for handling these problems by supporting the SME with functionality for on-line scoring of alternative plausible solutions and tagging of events for post-experimental investigation by addressing audio/video recordings. This technique also presents possibilities for the use of several SMEs and quality assurance by calculation of inter scorer reliability.

### **5.3.4 Measuring the plant performance**

A measure of plant performance is typically a measure of how well the plant is taken care of during normal operation or in the case of disturbances. This is done by an evaluation of the discrepancy between the "ideal" way of running the plant, defined by physical and mechanical criteria, and the way the plant actually behaved in the given situation.

## **5.4 Measurement problems**

### **5.4.1 Instructional effects on self-reporting**

In a part-task simulator-study of complex learning Stark, Gruber, Renkel and Mandl (1998) reported that self-evaluations were influenced by the learning context. Comparing a subjective measure and an objective measure, the subjective reporting was influenced by the amount of feedback given to them during the simulation. Subjects experiencing failure on actions, were likely to have a negative effect on the self evaluation of action knowledge, while subjects not getting feedback probably found it easier to fade out negative results and subjectively adjust the weighting or the relation between (experienced) success and failure in a favourable direction.

Some of the same symptoms were observed in a study of multiple fault detection scenarios (Follesø et al., 1995). In this study, novice and expert operators were compared on their ability to detect hidden malfunctions in a nuclear simulator study. Operators received no specific feedback on their problem solving, except the standard information given in the simulator interface. The novice operators with low detection rate, reported better results than the operators with better results on the detection task. The effect was explained by the interpretation that the former group was simply not aware of any second or third failure, they only knew the first one, while the latter group knew there were more failures in the system, but they were unable to find them. In Drøivoldsmo et al. (2002) the same problem was reported in a study of radiation awareness and situation awareness measured during testing of different modes for radiation presentation. In two different conditions of radiation presentation, subjects reporting high situation awareness, scored poorly on a test of radiation awareness and vice versa. In this study, the effect was explained by the inferior presentation of radiation in the first group preventing them from understanding the complexity of the situation.

The previously cited examples demonstrate that subjective reporting in complex tasks such as simulations is susceptible to distracting information. There is a long list of possible explanations for this distortion, ranging from cognitive phenomena such as bad memory for intentions versus actions, to social desirability, and self esteem. However, for some situations, admission to the information about the problem solving is most easily achieved through some type of self-reporting, and the advantages of the technique have to be evaluated against the possibly negative effects for each study depending of the purpose of the investigation.

### **5.4.2 Strengths and weaknesses of the SME**

As described in previous sections, a common way to rate performance in complex dynamic situations is to use a SME. Looking closer to this technique shows that it is not without problems. David Hume has been saying, in a practical interpretation, that everything the judge know about the topic he is judging has come to him from his senses. Hume argues that there is no substance. In other words, Hume says that what the SME observes are creations of his own imagination. In the situation where a SME is rating performance, we have to rely on the experience of this judge and his

interpretation of the situation. If we fully believed in Hume's view, and that knowledge shall be interpreted as the sum of consciously direct given experiences, this would mean a great variance from one SME to another in how to view correct solutions for how to solve problems in a complex dynamic situation. Luckily, we are able to judge and evaluate more general concepts derived from our knowledge and are not limited to only to the direct perception, but a number of factors influencing judgment can be identified.

To handle the influence from the general knowledge structures for understanding of problem and possible solutions (problem schemata) is important for good SME scoring. A schema is an underlying structure of knowledge for how we perceive the world and is used in the comprehension of situations. The schema is often viewed as a frame, which includes default assumptions and open slots for particular information (Medin and Ross, 1992).

As stated by Reason (1990):

*Systematic errors can arise (a) from fitting the data to the wrong schema, (b) from employing the correct schema too enthusiastically so that gaps in the stimulus configuration are filled with best guesses rather than available sensory data and (c) from relying too heavily upon active or salient schemata (p. 36)*

Considering the problem of the feeble human memory, the type of event to score could be critical for the SME judgment. Experts have a superior memory for patterns supported by their already well developed knowledge. Of particular interest in the results from studying chess players, is that experts understand and remember situations they study in which other people are performing the tasks. The expert's skill of remembering complex patterns is not limited to the expert's own actions, but they seem to limit to their own experienced situations and they do not generalise to other domains. For situations containing random information not matching patterns of experienced situations, or situations not logically derived from their expertise, the memory is on the same level as for a novice (Anderson, 1990; Glass and Holyoak, 1986; Parkin, 2000).

Procedural knowledge versus declarative knowledge can also influence the scoring. Procedural knowledge is knowledge whose existence is reflected in action and the stored information responsible for these actions cannot be verbalised (Parkin, 2000). Expert judgment and scoring depends on manifestation of the expert-judged outcome of a problem-solving process. The problem-solving process itself is often difficult to describe and this could limit some of the communication between SMEs and the other experimental staff planning, scoring, interpreting the outcome of studies.

Another important factor is the availability of the hypothesis. Easily available hypotheses will be over represented in the generation of possible solutions to the problem under study, and the easily available hypotheses will be based on previous experience (Parkin, 2000). In the case where the SME has to come up with a fast decision, the availability heuristic (Tversky and Kahneman, 1974) will lead the solutions towards solutions of the problem already known by the SME. Other, more specific sources of errors are the sequential effects, where the judgment of an item on the rating scale is affected by the preceding items, and intercorrelational error, where the rating of one characteristic affects ratings of other different characteristics (Meister, 1985).

With the absolute dependency on SMEs for almost all phases, outcome of an ecologically valid study will depend on how the problems and advantages of using the

SME are handled. Retrospective scoring based on the memory of the SME would from the current discussion not seem advisable. As discussed above (5.3.3), sufficient raw data resolution (continuous AV-recordings and simulator logs) together with a dedicated tool handling, the problems discussed can improve SME scoring considerably.

SME ratings in extended simulations, for example using VEs for the initial phases of integrated system validation could add additional problems. Since such studies would use systems not yet built, there is no way the SME is familiar with the system and additional evaluation of the SME capability should take place to ensure the correctness of the judgments. Another, mostly technical problem is to give the SME a full overview of the information provided to the subjects. In the case of subjects using HMDs for VR or AR information presentation, the HMD image has to be simultaneously displayed for the subject and the SME.

## **5.5 Data scoring and analysis**

### **5.5.1 Process tracing methods**

Complex and dynamic situations in experiments put new requirements on the tools for data scoring, aggregation and analysis. For simulations with professional operators, the cost of data gathering implies that one in most cases has to live with a limited number of subjects. Data scoring connected to the time line allows expanding the use of within subject design through use of time-windows based on important events or stages in the scenarios. Within-subject designs used across scenarios have also proved useful in the analysis in situations where effects are scenario dependent, although interpretation can be difficult when such explorative techniques are employed. The process tracing technique<sup>1</sup> is particularly useful for collecting data for this type of analysis. Even if the cost of scoring often can be higher compared to more discrete measurement (e.g., questionnaire data), the flexibility and better accuracy in defining critical events could justify this cost.

Data useful for more detailed design issues need a high resolution. The current problem of investigation of psychological variables in simulation is the same as for the human factors people using stand-alone experiments from the history of cognitive psychology: translation of behavioural conclusions into practical design input. To fill this gap the only way of progress is to get resolution of the resulting analysis to fit the demand from the designers (Meister, 1995).

A common goal for the data collection, scoring and analysis of data in HAMMLAB has been to make the performance scores comparable across scenarios, and sensitive to different levels of problem complexity. To achieve flexibility in the scoring, and to meet different demands for details, data have been recorded on the highest practical resolution and attached to a common time-line for all data sources. The types of data recorded have often been a combination of process tracing measures, different SME evaluations and questionnaire data. The high resolution also has the advantage of providing necessary data for post-hoc hypothesis and reanalysis of data from an experiment in light of new

---

<sup>1</sup> The term process tracing will be used as a generic term including all data collection that is continuously collected connected to a time line, extending Theureau's (2000) definition to cover e.g. simulator operator event logs and EMT.

knowledge. It is also useful for supplementary analysis when measurement can depend on several levels of analysis, e.g., single operator and team performance measurement.

Types of data suited for collection from an extended simulation would not be much different from the data collected in a simulation without VR/AR. Commonly used data relevant for the process tracing approach would be AV-data, EMT data, simulator logs in addition to SME ratings recorded with a sufficient frequency. In addition the VR/AR environment would require some control parameters. These are continuously logged data like framerate for the presentation of VR/AR, and calibration data from use of AR technology.

The type of data and technique for collection of data decides the amount of work required in the scoring and analysis phase. As discussed above (5.3.1), measures based on process tracing measures like verbal protocols are extremely time consuming. The same can be said for EMT used for scoring of "intentional information gathering" (Hauland, 2002). Since patterns of eye movements vary both within and between subjects, there is no effective automatic scoring of EMT available for this type of data. The most flexible and time efficient tools are semi computerised scoring tools (an overview of relevant tools is given in Sanderson and Fisher (1997)).

The main source of event-data from a simulator is the logs.

These can be divided into several categories; the logs of system parameters which make it possible to trace the fluctuation of flow, temperatures and pressure; the logs of actions performed by the operators or the experimental staff; and the logs of events from the autonomous control systems and alarm handling in the plant. The raw data from these logs can be collocated with the other process tracing data and aggregated into meaningful units as part of different indicators of human performance. One example of the usefulness of logs for the process tracing methods is clarification of verbal data. When important parts of information are unclear from the verbal data, meaning can be inferred from the logs (Hauland, 1996; Hollnagel et al., 1996; Kaarstad et al, 1995; Kaarstad, Kirwan, Follesø, Endestad and Torralba, 1994).

The traditional way of analysing experimental data is mostly based on analysis of variance (ANOVA). Funke (1995) argues that experimental studies of complex problem-solving suffer due to the misunderstanding that ANOVA is viewed as a research method and not taken for what it really is, a tool for data analysis. When it comes to testing a hypothesis, some kind of inferential statistic is necessary, and there are a number of these available. For the purpose of design evaluations and provision of data for the underfeed community of designers, this requirement is not absolute.





## **6 DISCUSSION OF THE EXTENDED SIMULATION**

As the technology develops and becomes commonly available, the use of VR/AR will most likely move into a number of everyday situations, and probably change what we today know as desktop computing and computer based process control. The reason for this is that these techniques allow interaction more like the interaction in the real world. The entertainment industry is a major driver of the technological progress in the computer industry and their target group is young people facing the new technology even in childhood. Some years ago university students were fascinated by a computer game in which they moved in one dimension on the computer screen by using the arrow keys on the keyboard: the object of the game was to kill space invaders. Today the computer game Quake can be played by using a head mounted display, mobile computer, head tracker, and GPS system to provide inputs to control the game (Piekarski and Thomas, 2002). The player of the game moves around the real world, and the monsters appear as though they were standing next to and behind real buildings. The technological development is inevitably affecting the way people act and think. For example, advanced mobile telephones and personal digital assistants (PDAs) are today common property. One could say that technology will (already has) augment the human problem solving by providing a number of aids, which act as natural extensions for everyday thinking.

This human behaviour augmented through computer interaction has implications for how the test situations should be in future experiments in the area of applied psychology. In the same way as the computerised test questionnaire is becoming the more "natural" test situation (Drøivoldsmo and Johnsen, 2002), the ecological setting could be some kind of VE.

### **6.1 *Limitations of the extended simulation***

The VR and AR worlds simulated with current technology are influenced by a number of potential problems. The most empirically explored and discussed concepts are constituted of a set of factors believed to influence the user's sense of being present in the VE.

"The laws of physics have bound reality, while the laws of psychology have bound virtual reality" (Chignell et al. 1999).

This observation implies that in the computerised worlds built by use of AR and VR technology, the limits for what can be presented are not restricted by physical laws and calls attention towards the knowledge of human perception, cognition and behaviour.

#### **6.1.1 SOP definition**

Witmer & Singer (1998) defined SOP as "the subjective experience of being in one place, even when one is physically situated in another," or as "experiencing the computer-generated environment rather than the actual physical locale" (p. 225). Regenbrecht et al. (1998) emphasise the cognitive construction of a space in which the

user's body and the VE seem to share the same world. However, as Stanney & Salvendy (1998) and Wilson (1997) point out, there is an absence of agreed methodology and measures for the SOP in a VE.

A fuller discussion of this definition, and the factors affecting SOP presented below, is given in Drøivoldsmo et al. (2001).

### **6.1.2 Factors affecting SOP**

The literature on SOP mentions the following factors as most important for producing a sense of presence: the ability to interact with the VE, user-initiated control, the extent and quality of sensory information, reduction of distracting elements in the physical environment, realism and meaningfulness of the VE, and field of view (Hoffman et al., 1998; Stanney and Salvendy, 1998; Wilson, 1997; Witmer and Singer, 1998). It is important that the user is immersed into the simulation so that the user has a sense of being present. The degree of immersion depends primarily on the hardware technology used, whereas presence is achieved by making an environment that causes the user to suspend disbelief and accept the computer-generated experience as 'real' (Louka, 1999).

The VE is presented to the viewer through a physical display, and it is important to be aware of the effects of different display characteristics. Two key characteristics are the frame rate and the display type, which may be head-mounted (HMD), desktop, or projected. There is probably a close relationship between perception and navigation in VEs. If perception (i.e. the ability to construct an accurate representation of the environment) in the VE is poor, then navigation is likely to be poor as a result.

Another important factor for successful use of VEs is to avoid simulator sickness. Simulator sickness is probably caused by a mismatch between the sensory stimuli felt by the body and the stimuli the organism expects based on past experience (Stanney and Hash, 1998). This mismatch cause illness that manifests itself in mental disorientation, nausea, drowsiness, fatigue, eyestrain and headache (Bullinger, Bauer and Braun, 1997).

At the current level of technological development, the factors discussed above constitute a number of limitations for what can be performed in research in VEs. However, with the current speed of innovation in the computer industry, technological deficit will probably constitute a smaller problem than the lack of knowledge about human behaviour.

## **6.2 *Extended control room simulation, artificial case example***

Among the lessons learned from the Three Mile Island accident was that the erroneous actions made by operators in a control room was a major contributing factor to the accident and its unsuccessful management. Among the factors disclosed as responsible for the accident and the operators unsuccessful management was a poorly designed control room not supporting the operators decision making, lacking provisions for the monitoring of the basic parameters of plant functioning (Moray and Huey, 1988).

The purpose of this example is to show how to establish an experimentally based research tool with emphasis on acquiring knowledge of human performance when

humans are interacting with a technological environment not yet implemented in the real world. The aim of such a tool is to improve the technical basis for developing efficient and safe solutions to the increasingly important issue of how, during the control room upgrade process, to best adapt complex technological systems to the human operator.

### **6.2.1 Upgrading hybrid control rooms**

The equipment in a process plant is always under modification. Upgrading plant and control-room equipment is a continual activity. Due to low flexibility of and unsure supply of spare parts, old conventional control panels are often replaced with computerised control systems or changed to newer conventional controls. The upgrade process is often stepwise, changing one system or sub-system at a time. As an undesired effect, the process of refurbishment and replacement often creates a situation with a number of different systems which lacks interface unity in the control room. One of the more frequent situations is the situation of gradually replacing conventional control panels with systems based on VDU process pictures creating a hybrid control room as discussed in chapter 3 above.

One of the major challenges in the hybrid control room upgrade process is to design the control room to support the operator's problem solving. Introducing a new HMI with a different allocation of controls between VDUs and conventional controls, introduces a new style of operation, and for some situations the need for information and communication between crew members can change. Theoretical studies are somewhat limited when investigating how operators adapt to the new interface in order to foresee all potential negative aspects and benefits. Hence, there is a need for experimental studies involving humans in the right technological environment.

### **6.2.2 A VR control environment for upgrading NPP turbine systems**

The example of semi-immersive VR from HAMMLAB described in chapter 4.3.2 could be realised by combining VR screens and large overview screens connected to the simulator. This solution requires a detailed VR model of the stand-up control panels from the plant simulated, with relevant controls and meters modelled, and connected to the simulator. It also requires a control mechanism for the turbine operator to perform control actions in the VE, e.g., when operations in the VE are to be performed, the turbine operator is standing up and walking to a workplace where he can control the virtual operations. These operations are as shown in Figure 12 performed by a manikin. Alternatively, the operator could perform control actions in the VE by using data-gloves for direct manipulation. In this way, actions such as walking to the panels, starting/stopping systems, acknowledging alarms and reading values from the panels can be performed.

The process of moving systems from the control panels to the VDUs should take care of the crew's need for shared information about the turbine system. In our example, this is taken care of by moving the information in question from the turbine VR model and adding it to the large screen overview display as shown in Figure 13.



*Figure 12. Manipulated picture from HAMMLAB where the turbine operator (to the right) works together with a manikin reactor operator in the situations where operation and monitoring of control panels are necessary.*

Interaction between the process control simulator and the VE is described in Karlsson, Drøivoldsmo, Johnsen and Kvalem (2002). The main principle is that the system will consist of two connected VR models that allow for team communication in the hybrid / virtual control room. The first will be projecting to a large back-projected screen the VR model of the control room as seen from the fixed position of the turbine operator's workplace. This fixed view will always show the reactor panels and all manikin activities in the virtual control room in front of those panels. A typical movement in this VR model will be the virtual reactor or turbine operator performing operations on the panels. While sitting by the turbine desk, the turbine operator can virtually see all movements of the reactor operator on the large screen and vice versa.

The second VR model will be used by the operators to simulate work operations such as retrieval of documents and operations on the conventional VR panels. When virtual operations are to be performed, the reactor operator is standing up (illustrated in Figure 7) and walking to a workplace where VR operations are performed. This workplace is not illustrated in Figure 13. Sitting down on this workplace, the operators start moving in the VR model that is displayed on the monitor. Now actions like walking to the panels, acknowledging alarms and reading values can be performed. While these actions are done, all operator movements are illustrated by a manikin. These actions will be visualised in the operators' view.

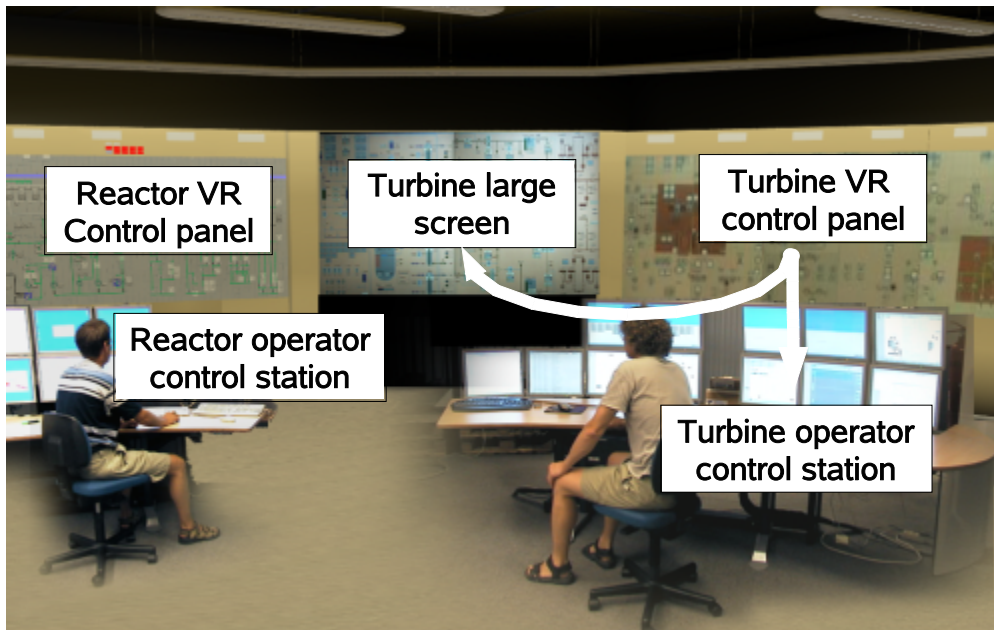


Figure 13. Moving systems and subsystems from the turbine VR control panels to the operator control station and large screen.

### 6.3 VR as a tool for simulation of field operations in control room studies

The task of running the plant requires a considerable collaboration between the control room operators and the field operators. Current simulations do normally not account for realistic co-operation between control room and field operators. Outside control room activities are simulated by SME role-play and estimation of time for the field operator carrying out the work operators requested from the control room. By calling the SME the control room operators can get in contact with a field operator and request manual inspection, or operation of equipment. The SME playing the role of the field operator is then replying (with a best guess estimated time delay accounting for the time of, e.g., manually closing a valve) on the telephone request and informing about status of equipment and the results of operations. Obviously, working with humans, there are situations where the current handling of outside control room actions can influence the outcome of the simulation. By using VEs, simulation and data collection, the functions outside of the control room can be taken into account in the study of co-operation in existing systems, as well as work with new systems not yet built. In particular, the verification and validation of new technology available for the field operator, and test methodology needed to investigate the effect of introducing this technology into the NPP.

Would VR be the best tool for a simulation of outside control room activity or are there other more efficient ways of achieving equally good results? The ideal situation for studies of field operation would be to communicate with field operators in the plant under study. However, such set-ups are often not feasible due to practical constraints. Access to the areas or equipment under study is not always granted, and critical operations cannot be performed on running equipment. On the other hand, connecting a

VE to a simulator will allow for most types of emergency or maintenance operations without restricting the content of scenarios.



*Figure 14. Example of VE used for field operation. Field operation simulation in VR gives the opportunity to simulate situations where the field operator is prevented from doing his job by physical or organisation constraints. The person wearing the HMD can see his colleague in the virtual environment.*

Relevant situations for using a technique like this would be when more extensive (realistic) communication with the operators have impact on the quality of the work the operators have to handle during a scenario. Using VEs also give the opportunity to test collaboration between several field operators (see Figure 14), a situation hard to create by role-play without any aiding context in the simulator gallery. Other applications of this approach would typically be in the early phases of the design process. Task analysis, human reliability analysis and training are also areas in need for support tools.

The findings from the control room experiments in this thesis, Drøivoldsmo et al. (2001) and Nystad, Helgar et al. (2002) testing the effectiveness of VEs, indicate that using VEs for real world tasks should be supported by dedicated tools for situations requiring perceptually based problem solving, and that the navigation aids in the model are crucial for the subject's performance in the VE. In the VR and AR field operation experiments, the findings indicate that the VEs has a positive effect on spacial learning, supporting the hypothesis that VEs could be used for the purpose of collection human performance data in such simulations. An important observation from the two field operation experiments, reported in Drøivoldsmo and Johnsen (2002) was the impact of environmental cues from the test surroundings. Human performance measurement in these extended simulations has a need for flexible and efficient data collection tools providing means for immediate testing well adopted to the test situation. This example has mostly considered VR as the tool for extended simulation, but there is no reason to think AR should be less efficient given that access to relevant test areas can be provided.

#### ***6.4 Final comments on psychological experiments in VR***

Both in basic and applied experimental psychology there has been a shift from a smaller and simpler paradigm to the analysis of complete tasks and skills. Traditionally the area of basic psychological research has been occupied with the description of mental processes involved in human performance, while engineering psychology has focused on how man and machine have worked together. One way of distinguishing between these two domains has been to look at engineering psychology as a bridge between the basic psychological knowledge (sensory, perceptual, motor and cognitive capabilities) and the design of human systems (Nickerson, 1999). Depending on the future of experimental psychology, the gap between experimental psychology and the practical implications of the research may be closing. However, history has shown that a high level theory of practice is difficult to obtain and that the link between scientific value and relevance to practical application is uncertain (Pylyshyn, 1991). Some of the advancement must be made in the area of a better foundation for understanding the implications of using psychological methodology in the applied areas and establishment of reoccurring patterns that are related to their function. Since technology is influencing psychology, one could also hope that an effort is being made to specialise the psychological science for areas relevant for the HMI design process.

Psychologists have always been attuned to social environments and available experimental tools (Chignell et al. 1999). The availability of the tachistoscope, Skinner box and laboratory rat heavily influenced the direction of psychological research. For psychology in general, the technique of controlled presentation of virtual complex dynamic scenes, will allow experiments a higher degree of ecological validity in a number of new settings. High ecological validity has traditionally been seen as an impediment to the generalizability of results (Banaji and Crowder, 1989). Accepting VEs as ecologically valid and making use of virtual scenes in the university laboratories allows for a new type of studies, keeping a high number of participants in experimental research. The limits for use of VEs and the speed of development is impossible to foresee, but current technology already allows multiple agents acting together in virtual societies.





## 7 CONCLUSION

The future of engineering psychology is of great importance to the insight in human problem-solving and establishment of knowledge in areas where humans are in control of highly complex information systems. Before the Three Mile Island accident, a NPP was primarily perceived from the technical perspective. After the Three Mile Island, this view was enlarged to cover the human operation in the control room and focus was set on the man-machine interface. This shift in focus has been a first step towards more safe and reliable operation, and the next step could be to include all operators running the plant, with focus on the co-operation between the crew in the control room and the operators outside the control room. The new techniques for simulation described here have the potential to provide more realistic situations including the critical aspects of a field operators manual interaction with the process.

Under the assumption that a simulator can recreate the process it is intended to represent, and that the subjects used for testing have expertise in the special field, keeping realism with a higher level experimental control, seem useful for the purpose of providing data for building safer high-risk systems. The higher level experimental control relies on the trained subject's approximately constant behaviour in realistic situations. The high cost of such experiments can be reduced by more effective data collection procedures (Drøivoldsmo and Johnsen, 2002) and more careful design of the experiments with a view to data reuse.

The results from the VR control room design verification experiments (Drøivoldsmo et al. 2001; Nystad, Helgar et al. 2002) indicate that satisfactory use of a VE can take place only if the VE provides dedicated support for the user tasks. A VE usable for the purpose of realistic studies has to seem real to the user, with a smooth look and good light settings. The users immersion into the VE is crucial to create the "right" feeling and touch to provide SOP. The experiments partially succeeded to do this using the less immersive desktop VR technique. The two field operator experiments (Nystad, Drøivoldsmo et al. 2002; Drøivoldsmo et. al 2002) both showed successful in using VR and AR for realistic, real world tasks. In this respect the experimental part of this thesis has been successful in providing insights in the basic technological and methodological needs for using the new technology in human performance experimentation.

Refurbishment and building of new control rooms can benefit from using the VR technology through the whole design phase (Drøivoldsmo and Louka, 2002).

The examples given in chapter 6 have shown how an experimentally based research tool could be built by combining the current and new technology. The new technology, if successfully implemented, could provide major benefits in the human interface design process allowing predicting human performance before the systems are operational, indeed early in the conceptual and detailed design stages.

This thesis is meant as a step towards a framework for methodological development using new technology in human performance simulator experiments. The utility of the framework depends on the extent of its applicability to a much wider range of problems than has been considered here, but the principles for experimentation, data collection and analysis demonstrated would seem transferable to most areas considering control of

physical equipment and may in the future be considered as a prerequisite for performing applied psychological research.

## 8 REFERENCES

- American Institute of Aeronautics and Astronautics. (1992). *Guide to human performance measurement* (BSR/AIAA G-035-1992). Washington DC: American institute of Aeronautics and Astronautics.
- Anderson, J.R. (1990). *Cognitive psychology and its implications*. New York: W. H. Freeman and Company.
- Andresen, G. & Drøivoldsmo, A. (2000). *Human performance assessment: Methods and Measures* (HPR-353). Halden, Norway: OECD Halden Reactor Project.
- Azuma, R. (1997). A Survey of Augmented Reality. *Presence*, 6, 355-385.
- Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S. et al. (2001). Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, 21, 34-47.
- Banaji, M. and Crowder, G.C. (1989). The Bankruptcy of Everyday Memory. *American Psychologist*, 44, 1185-1193.
- Bargiela, A. (2000). *Strategic Directions in Simulation and Modelling*. Nottingham UK: The Nottingham Trent University Computing Department.
- Brehmer, B. (1995). Feedback delays in complex dynamic decision tasks. In P.A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Brooks, F.P. (1996). The Computer Scientist as Toolsmith II. *CACM*, 39, 66-69.
- Braarud, P.Ø. & Nystad, E. (2002). *Benchmark Validation of the upgraded Oskarshamn 1 Control Room* (IFE/HR/F-2002/1197). Halden, Norway: Institutt for Energiteknikk.
- Braarud, P.Ø., Drøivoldsmo, A. & Hollnagel, E. (1997). *Human error analysis project (HEAP) - the fourth pilot study: Verbal data for analysis of operator performance* (HWR-495). Halden, Norway: OECD Halden Reactor Project.
- Bullinger, H.J., Bauer, W. & Braun, M. (1997). Virtual environments. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1725-1759). NY: John Wiley and Sons, INC.
- Cacciabue, P.C. (1998). *Modelling and simulation of human behaviour in system control*. London: Springer Verlag.
- Campbell, D.T. and Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carroll, J.M. (1991). Introduction: The kittle house manifesto. In J.M. Carroll (Ed.), *Psychology at the human - computer interface* (pp. 1-16). New York: Cambridge University Press.
- Chalmers, A.F. (1999). *What is this thing called science?* (3rd ed.). London: Open University Press.

- Chignell, M., Hancock, P.A. & Takeshita, H. (1999). Human - computer interaction: The psychology of augmented human behaviour. In P.A. Hancock (Ed.), *Human performance ergonomics* (pp. 291-328). San Diego, CA: Academic Press.
- Cohen, R.J. & Swerdlik, M.E. (2001). *Psychological testing and assessment: An introduction to test and measurement*. (5th ed.). New York: McGraw-Hill.
- Coombs, C.H. (1964). *A theory of data*. New York: John Wiley and Sons.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-Experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Cronbach, L.J. (1990). *Essentials of psychological testing*. (5th ed.). New York: HarperCollinsPublishers.
- Drøivoldsmo, A. (2000). Study of errors by means of simulation and training. In P.F. Elzer, R.H. Kluwe & B. Boussoffara (Eds.), *Human error and system design and management* (pp. 201-206). London: Springer-Verlag.
- Drøivoldsmo, A. (2002). *The Halden Questionnaire help pages [On-line]*. Available: <http://psyweb.svt.ntnu.no/asgeird>: Internet.
- Drøivoldsmo, A. and Johnsen, T. (2002). Development and testing of an Internet-based data collection technique for simulator and real world experimentation. Manuscript submitted for publication.
- Drøivoldsmo, A., Helgar, S., Nystad, E., Johnsen, T., Louka, M.N. & Seregi, L. (2000). Verification of workplace design guidelines for the control room: Can virtual reality replace the physical mock-up? In K.I. Fostervold & T. Endestad (Eds.), *Proceedings of Nordiska Ergonomisällskapet's Årskonferanse, 2000* (pp. 73-77). Stockholm, Sverige: Nordiska Ergonomisällskapet.
- Drøivoldsmo, A. & Louka, M.N. (2002). Virtual reality tools for testing control room concepts. In B.G. Lipták (Ed.), *Instrument Engineers' Handbook, Vol.3* (pp. 299-306). Boca Raton, FL: CRC PRESS LLC.
- Drøivoldsmo, A., Nystad, E. & Helgar, S. (2001). *Virtual reality verification of workplace design guidelines for the process plant control room* (HWR-633). Halden, Norway: OECD Halden Reactor Project.
- Drøivoldsmo, A., Reigstad, M., Shimoda, H., Louka, M.N., Helgar, S., Gustavsen, M. & Nystad, E. (2002). *Augmented reality for visualisation of radiation* (HWR-717). Halden, Norway: OECD Halden Reactor Project.
- Drøivoldsmo, A., Skraaning, G., Sverrbo, M., Dalen, J., Grimstad, T. & Andresen, G. (1998). *Continuous measures of situation awareness and workload* (HWR-539). Halden, Norway: OECD Halden Reactor Project.
- EATMP Human Resources Team. (2000). *Simulations Facilities for Air Traffic Control Training* (HUM.ET1.ST07.3000-REP-02). Brussels: Eurocontrol.
- Endsley, M.R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37, 32-64.

- Endsley, M.R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M.R. Endsley & D.J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147-173). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ericsson, K.A. and Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Ericsson, K.A. & Simon, H.A. (1993). *Protocol Analysis: Verbal reports as data, Revised Edition*. Cambridge, MA: MIT Press.
- Federico, P.-A. (1995). Expert and Novice Recognition of Similar Situations. *Human Factors*, 37, 105-122.
- Follesø, K., Drøivoldsmo, A., Kaarstad, M., Collier, S. & Kirwan, B. (1995). *Human error - the third pilot study* (HWR-430). Halden, Norway: OECD Halden Reactor Project.
- Follesø, K., Kaarstad, M., Drøivoldsmo, A. & Kirwan, B. (1995). Relations between task complexity, diagnostic strategies and performance in diagnosing process disturbances. *European Conference on Cognitive science, Approaches to process control*, 5. Espoo, 1995-08-30.
- Fracker, M.L. (1991). *Measure of Situation Awareness: Review and Future Directions* (AD-A262 672). Springfield, Virginia, US.: National Technical Information Service.
- Funke, J. (1995). Experimental research on complex problem solving. In P.A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 243-268). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Geddes, B.J. (2002). Upgrading the control room. In B.G. Lipták (Ed.), *Instrument Engineers' Handbook*, Vol.3 (pp. 307-322). Boca Raton, FL: CRC PRESS LLC.
- Glass, A.L. & Holyoak, K.J. (1986). *Cognition*. New York: Random House.
- Gravelle, M.D. & Schopper, A.W. (1993). *Measures and methods for characterizing operator performance in nuclear power plant control rooms* (CSERIAC-RA-93-007). Wright-Patterson Air Force Base, OH: CSERIAC.
- Haga, Å. (2002). Georg W. F. Hegel. In T. Berg Eriksen (Ed.), *Vestens store tenkere* (pp. 249-264). Oslo, Norway: Aschehoug.
- Han, S.H., Williges, B.H. and Williges, R.C. (1997). A paradigm for sequential experimentation. *Ergonomics*, 40, 737-760.
- Haraldsen, A. (1999). *Den forunderlige reisen gjennom datahistorien*. Oslo, Norway: Tano Aschehoug.
- Hauland, G. (1996). *Building a methodology for studying cognition in process control: a semantic analysis of visual and verbal behaviour. Post graduate thesis in psychology*. Trondheim, Norway: Norwegian University of Technology and Science.
- Hauland, G. (2002). *Measuring team situation awareness in training of en route air traffic control: Process oriented measures for experimental studies (Ph.D. thesis)*. Aarhus, Denmark: University of Aarhus.
- Heim, M. (1998). *Virtual realism*. New York: Oxford University Press.

- Hellander, M.G. (1997). The human factors profession. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 3-16). NY: John Wiley and Sons, INC.
- Hoffman, H.G., Prothero, J., Wells, M.J. and Groen, J. (1998). Virtual chess: Meaning enhances user's sense of presence in virtual environments. *International Journal of Human-Computer Interaction*, 10, 251-263.
- Hollnagel, E., Braarud, P.Ø., Drøivoldsmo, A., Follesø, K., Helgar, S. & Kaarstad, M. (1996). *Human error analysis project (HEAP) - the fourth pilot study: Scoring and analysis of raw data types* (HWR-460). Halden, Norway: OECD Halden Reactor Project.
- Howell, W.C. (1993). Engineering psychology in a changing world. *Annual Review of Psychology*, 44, 231-263.
- Humbel, C. (2002). Validation of advanced human-system interfaces. An example of a human factors engineering process in Switzerland. *Proceedings of Japan-Halden MMS workshop* (pp. 50-53). Kyoto, Japan: Symbio Community Forum, Graduate School of Energy Science, Kyoto University.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: The MIT Press.
- IEC 964. (1989). *Design for control rooms of nuclear power plants* (International Standard 964 (1989-03). Geneva: International Electrotechnical Commission.
- ISO 11064-1:2000. (2000). *Ergonomic design of control centres - Part 1: Principles for the design of control centres* (ISO 11064-1:2000). Brussels: Belgium: ISO.
- Jokstad, H. & Sundling, C.V. (2002). Picasso-3 User Interface Management System. *Proceedings of the man-machine systems research sessions at the enlarged Halden programme group meeting* (volume 1) (pp. C3 P9). Halden, Norway: OECD Halden Reactor Project.
- Jorna, P. (2000). Context simulation: An interactive methodology for user-centered system design and future operator behavior validation. In N.B. Sarter & R. Amalberti (Eds.), *Cognitive engineering in the aviation domain* (pp. 181-209). NJ: Lawrence Erlbaum Associates.
- Karlsson, T., Drøivoldsmo, A., Johnsen, T. & Kvaalem, J. (2002). *Hybrid Control Room Studies: Integrating virtual reality (VR) technology and overview displays with the HAMBO simulator (project proposal for the 4 Swedish nuclear power stations)*. Halden, Norway: Institutt for Energiteknikk.
- Kerlinger, F.N. (1986). *Foundations of behavioural research* (3rd ed.). Orlando, FL: Harcourt Brace College Publishers.
- Kirk, R.E. (1995). *Experimental design: Procedures for the behavioural sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Kaarstad, M., Follesø, K., Collier, S., Hauland, G. & Kirwan, B. (1995). *Human error - the second pilot study* (HWR-421). Halden, Norway: OECD Halden Reactor Project.
- Kaarstad, M., Kirwan, B., Follesø, K., Endestad, T. & Torralba, B. (1994). *Human error - the first pilot study* (HWR-417). Halden, Norway: OECD Halden Reactor Project.

- Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge, UK: Cambridge university press.
- Lewin, M. (1987). *Understanding psychological research*. Malabar, FL: Robert E. Krieger Publishing Co.
- Louka, M.N. (1999). *An introduction to virtual reality technology* (HWR-588). Halden, Norway: OECD Halden Reactor Project.
- Lübce, P. (1983). *Filsofileksikon*. Saltrød: Forsythia.
- Medin, D.L. & Ross, B.H. (1992). *Cognitive Psychology*. Orlando, FL: Harcourt Brace Jovanovich.
- Meister, D. (1985). *Behavioral analysis and measurement methods*. New York: John Wiley & Sons.
- Meister, D. (1995). Experimental Analysis and Measurement of Situation Awareness: A Commentary. In D.J. Garland & M. Endsley (Eds.), *Experimental Analysis and Measurement of Situation Awareness* (pp. 43-47). Daytona Beach, Florida USA: Embry-Riddle Aeronautical University Press.
- Milgram, P and Kishino, F. (1994). A Taxonomy of Mixed reality Visual Displays. *IEICE Transaction of Information Systems*, 12, 1321-1329.
- Moray, N. (1997). Human factors in process control. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1944-1971). NY: John Wiley and Sons, INC.
- Moray, N.P. & Huey, B.M. (1988). *Human factors research and nuclear safety*. Washington, D.C.: National academy press.
- Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.
- Netland, K. & Øyen Hol, J. (1977). *Development of a Computer and Colour Display Based Control Room. (OPCOM) For Experimental Operation of the Halden Reactor* (HPR-202). Halden, Norway: OECD Halden Reactor Project.
- Nickerson, R.S. (1999). Engineering psychology and ergonomics. In P.A. Hancock (Ed.), *Human performance Ergonomics* (pp. 1-45). San Diego, CA: Academic Press.
- Nisbett, R.E. and Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Nunnally, J.S. & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- NUREG-0711. (2002). *U.S. Nuclear Regulatory Commission. Human Factors Engineering Program Review Model* (NUREG-0711, Rev. 1). Washington, DC, USA: The NRC Public Document Room.
- Nystad, E., Drøivoldsmo, A. & Sebok, A. (2002). *Use of radiation maps in a virtual training environment for NPP field operators* (HWR-681). Halden, Norway: OECD Halden Reactor Project.
- Nystad, E., Helgar, S. & Drøivoldsmo, A. (2002). *Virtual reality verification of workplace design guidelines: A follow-up study* (HWR-718). Halden, Norway: OECD Halden Reactor Project.

- O'Hara, J.M. (1999). A quasi-experimental model of complex human system validation. *Cognition, Technology & Work*, 1(1), 37-46.
- O'Hara, J., Stubler, W., Higgins, J. & Brown, W. (1997). *Integrated system validation: Methodology and review criteria* (NUREG/CR-6393). Washington, DC: US Nuclear Regulatory Commission.
- Papin, B. (2002). Integration of human factors requirements in the design of future plants. *Proceedings of the man-machine systems research sessions at the enlarged Halden programme group meeting (volume 1)* (pp. C3 P1). Halden, Norway: OECD Halden Reactor Project.
- Parkin, A.J. (2000). *Essential Cognitive Psychology*. Philadelphia, PA: Psychology Press Taylor & Francis.
- Piekarski, W. and Thomas, B. (2002). ARQuake: The Outdoor Augmented Reality Gaming System. *Communications of the ACM*, 45, 36-38.
- Pylyshyn, Z.W. (1991). Some remarks on the theory-practice gap. In J.M. Carroll (Ed.), *Psychology at the human - computer interface* (pp. 39-49). New York: Cambridge University Press.
- Reason, J. (1990). *Human Error*. Cambridge, UK: Cambridge University Press.
- Regenbrecht, H.T., Schubert, T.W. and Friedmann, F. (1998). Measuring the sense of presence and its relations to fear of heights in virtual environments. *International Journal of Human-Computer Interaction*, 10, 233-249.
- Reither, F. (1981). About thinking and acting of experts in complex situations. *Simulation & Games*, 12, 125-140.
- Rogovin, M. & Frampton, G.T. (1980). *Three Mile Island: A report to the commissioners and the public* (vol 1). Springfield, VA: National Technical Information Services.
- Roth, E. & O'Hara, J. (1998). *Integrating digital and conventional HSIs: Lessons learned from a control room modernization program* (BNL letter report J6012-T3-4-5/98). Upton, NY: Brookhaven National Laboratory.
- Rouse, W.B. (1991). *Design for success: A human centred approach to designing successful products and systems*. New York: John Wiley & Sons.
- Sanderson, P. & Fisher, C. (1997). Exploratory sequential data analysis: Qualitative and quantitative handling of continuous observational data. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1471-1513). NY: John Wiley and Sons, INC.
- Sebok, A., Nystad, E. & Drøivoldsmo, A. (2002). Improving safety and human performance in maintenance and outage planning through virtual reality-based training systems. In J.J. Persensky, B. Hallbert & H. Blackman (Eds.), *New century, new trends: Proceedings of the 2002 IEEE 7th conference on human factors and power plants* (pp. 814-822). New York: IEEE.
- Sanders, A.F. (1991). Simulation as a tool in the measurement of human performance. *Ergonomics*, 34, 995-1025.



- Sheridan, T.B. (1999). Descartes, Heidegger, Gibson, and God: Toward an eclectic ontology of presence. *Presence*, 8, 551-559.
- Shimoda, H., Ozawa, T., Samejima, R. & Yoshikawa, H. (2002). An experimental study of a training system for diagnosing NPP anomaly using eye-sensing HMD. *Proceedings of Japan-Halden MMS workshop, March 4-5, 2002* (pp. 86-90). Kyoto, Japan: Symbio Community Forum, Graduate School of Energy Science, Kyoto University.
- Simon, C.W. (1973). *Methods for handling sequence effects in human factors engineering experiments*. (Technical report no. P74-415 A). Culver city, CA: Huges Aircraft Company.
- Skourup, C. and Reigstad, M. (2002). Operators in process industry: Interaction with wearable computers. *Cognition, Technology & Work*, 4 (4), 245-255.
- Skraaning, G. (1998). *The operator performance assessment system (OPAS) (HWR-538)*. Halden, Norway: OECD Halden Reactor Project.
- Skraaning, G. (2002). *Realism versus experimental control: Methodological challenges for simulation studies in complex operating environments (Working report)*
- Stanney, K.M. and Hash, P. (1998). Locus of user-initiated control in virtual environments: Influences on cybersickness. *Presence*, 7, 447-459.
- Stanney, K. and Salvendy, G. (1998). Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction*, 10, 135-187.
- Stark, R., Gruber, H., Renkl, A. and Mandl, H. (1998). Instructional effects in complex learning: Do objective and subjective learning outcomes converge. *Learning and Instruction*, 8, 117-129.
- Theureau, J. (2000). Nuclear Reactor Control Room Simulators: Human Factors Research and Development. *Cognition, Technology & Work*, 2(2), 97-105.
- Tversky, A. and Kahneman, D. (1974). Judgments under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Wann, J. and Mon-Williams, M. (1996). What does virtual reality NEED?: Human factors issues in the design of three-dimensional computer environments. *Int j hum-comp stud*, 44, 829-847.
- Wickens, C.D. & Hollands, J.G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wilson, J.R. (1997). Virtual environments and ergonomics: needs and opportunities. *Ergonomics*, 40, 1057-1077.
- Witmer, B.G. and Singer, M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7, 225-240.

## **8.1 List of acronyms and abbreviations**

ANOVA	Analysis of Variance
AR	Augmented Reality
ATC	Air Traffic Control
CAD	Computer Assisted Drawing
COSS	Computerised Operator Support System
CVP	Concurrent Verbal Protocol
EATMP	European Air Traffic Management Programme
EMT	Eye Movement Tracking
GPS	Global Positioning System
HAMBO	HAMMLAB boiling water reactor
HAMMLAB	Halden Man-Machine Laboratory
HEAP	Human Error Analysis Project
HMD	Head-Mounted Display
HMI	Human - Machine Interface
HVRC	Halden Virtual Reality Laboratory
HWR	Halden Work Report
NPP	Nuclear Power Plant
PC	Personal Computer
SIMULA	SIMUlation LAnguage
SME	Subject Matter Expert
SOP	Sense of Presence
S-R	Stimulus - Response
TMI	Three Mile Island
VDU	Video Display Unit
VE	Virtual Environment
VR	Virtual Reality
WWW	World Wide Web

# Paper A

Drøivoldsmo, A. and Johnsen, T. (2002). Development and testing of an Internet-based data collection technique for simulator and real world experimentation. Manuscript submitted for publication.

Copyright permissions are not arranged for electronic copies of this paper. Contact the authors for more information



# Paper B

Drøivoldsmo, A. & Louka, M.N. (2002) Virtual reality tools for testing control room concepts. In B.G. Lipták (Ed.), *Instrument Engineers' Handbook, Vol.3* (pp. 299-306). Boca Raton, FL: CRC PRESS LLC.

Copyright permissions are not arranged for electronic copies of this chapter. The work can be found in the Instrument Engineers Handbook.



# Paper C

Drøivoldsmo, A., Reigstad, M., Shimoda, H., Louka, M.N., Helgar, S., Gustavsen, M. & Nystad, E. (2002). *Augmented reality for visualisation of radiation (HWR-717)*. Halden, Norway: OECD Halden Reactor Project.





# Use of Radiation Maps in Augmented Reality

(HWR-717)

Asgeir Drøivoldsmo, Magnus Reigstad\*, Hiroshi Shimoda\*\*, Michael N. Louka, Stein Helgar, Morten Gustavsen and Espen Nystad

\* Norwegian University of Technology and Science (NTNU), Dept. of Engineering Cybernetics, 7491, Trondheim, Norway

\*\* Kyoto University, Graduate School of Energy Science, Gokasho, Uji, Kyoto 611-0011, Japan

## Abstract

The entrance of small wireless computers opens for application of computers in a number of new areas. In the near future, operators will be able to wear a computer much in the same way as eyeglasses or clothing are worn today. One example of a potentially useful application of this technology is visualisation of radiation. With the augmented reality (AR) technique, combining computer-generated pictures with the real world, information of variation in radiation levels can be presented for field operator working in contaminated areas.

Using the AR system prototype developed for the experiment described in this report, information about radiation was presented to the subjects using different sensing modalities. The visual mode was chosen as the main mode and compared to a combination of visual, auditory and haptic presentations. A control group presented with paper-based information about levels of radiation performed the same task as the groups using the AR prototype.

The results indicate that the operators' awareness of radiation was the same after using the AR prototype as in the control group. The auditory and haptic information presented seemed to have a positive effect on the operators' radiation awareness.

The report discusses the results from the experimental test and from a usability test of the prototype. Based on these results and current technology status in the domain of AR, recommendations for further development are given.

# 1. ACKNOWLEDGEMENTS

This experiment is the fourth experiment in a series of human factors experiments testing the application of virtual reality (VR) technology to real-world tasks. Development of the AR system used in this experiment has been done as an extension of the 3D modelling and development work done for the VR-training experiment reported in Nystad, Drøivoldsmo and Sebok (2002). The same project team with members from the VR, Reactor operation, Radiation protection and Psychology divisions at the HRP was participating in the development and planning of the experiment. The same group has participated through the specification, design, prototyping and evaluation phases in all four experiments. We would like to express our gratitude to the project team members, in particular Thomas Elisenberg, Geir Mjønes, Heidi Andersen and Erik Skattum who made this project possible with their co-operative attitude, broad operational experience and insightful help during the whole project. Niels-Kristian Mark, who helped us build the VR-model of the reactor hall, and Terje Johnsen who developed the human performance data collection system used in the experiment.

Development and implementation of the sound and vibrations module used in the experiment was performed at the Department of Engineering Cybernetics at the Norwegian University of Science and Technology. Special thanks go to Asle Olsen who did a great job in the development of the code for the sound and vibrations module.

We would also like to thank all subjects who participated, as without their contribution this project could not have been carried out.

## **2. INTRODUCTION**

The limits for human radiation exposure set by the International Commission on Radiation Protection (ICRP) are 50 milli-Sievert (mSv) a year, or maximum 100 mSv exposure during a five year period. Nuclear facilities have a comprehensive control of their workers and are constantly trying to find ways to reduce workers radiation doses. This work should continue with all available funds, and new technology should be tested, implemented and made available to the operators.

Radiation in contaminated areas varies from one location to another within the overall environment of interest. One way of reducing radiation doses is to provide better information to the operators in contaminated areas about levels of radiation in their immediate surroundings. Given that all physical shielding efforts have been carried out, there are two ways to reduce radiation doses received by an operator. Reducing the time spent near radioactive sources or moving further away from them will both lead to decreased exposure. Doubling the distance from the operator's position to the radioactive source will reduce the dose by the square root of the original dose, and halving the time will reduce the dose by 50%.

A significant problem with gamma rays is that the human eye cannot see them, and so one way to achieve the goal of giving operators better information about varying levels of radiation would be to "make them visible" and thus enable operators to spend more time in low radiation areas. The scope of the project reported here has been to develop a prototype system that enables operators to see otherwise invisible radiation, and to test it to examine how such a system should be developed to support human senses in the best possible manner. The purpose of the system was to give individual process plant field operators information updated in real-time about the radiation surrounding them, based on their actual location in the real environment and on their direction of sight. The information was presented to the individual operator via a see-through head-mounted display (HMD) and stereo audio signals and physical vibrations.

The report discusses the use of AR for industrial purposes and provides an overview of different application areas. Further, the report introduces basic research, and dominating views on attention and multi-modal perception, before presenting an experimental test of the system, where it is used by operators working in contaminated areas. A schematic description of the system is given in the method chapter of the report, and recommendations are given for further testing and development of AR technology in the Halden Project based on the results of the experiment.

## **3. AUGMENTED REALITY**

### **3.1 What is Augmented Reality?**

Azuma (Azuma, 2001) defines an AR system as one that "supplements the real world with virtual (computer-generated) objects that appear to coexist in the same space as the real world". Wearing a HMD, for example, the user of an AR system can see the real world around him, but with additional virtual objects appropriately aligned (or "registered", in AR terminology) with real objects. As the user moves around in the real

environment the virtual objects maintain correct alignment with the physical environment resulting in a seamless combination of real and virtual objects. AR can thus enhance the user's perception of the real world.

When defining AR, Azuma lists the following three properties that an AR system should typically exhibit:

- combines real and virtual objects in a real environment
- runs interactively, and in real time; and
- registers (aligns) real and virtual objects with each other.

This definition is not restricted to visual information and is equally applicable to the other senses, however most existing AR systems place an emphasis on the display of visual information. In the mobile AR system used in the experiment described later in this report, visual, auditory, and tactile information are combined.

AR is often viewed by researchers (e.g. Azuma, 1997) as a variation of VR and the two concepts are sometimes confused, however VR and AR are conceptually very different. While VR technology is used to immerse the user in a virtual environment, replacing the real world with an entirely computer-generated one, AR seeks to supplement the real world, in which the user is already immersed, with additional information. Milgram (Milgram, 1994) describes AR as a mixed reality, placed between the real environment and a virtual environment. Milgram's reality-virtuality continuum is shown in Figure 1. Augmented virtuality is another type of mixed reality, where real objects are added to virtual ones, as opposed to augmented reality where virtual objects are added to real ones.

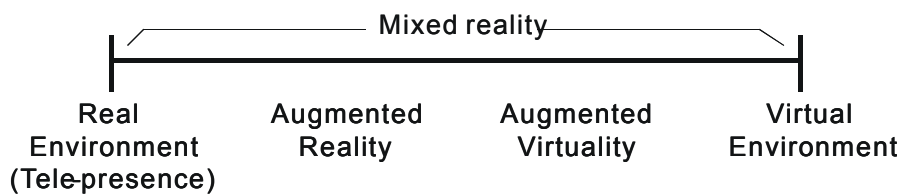


Figure 1. Milgram's reality-virtuality continuum.

Sutherland laid the foundations of AR research in the 1960s, however AR is relatively young as a research field in its own right, with the bulk of the research work done during the last decade.

Researchers are pursuing the values and possibilities of AR. The essential value of AR is that it can enhance human perception and extend conventional interaction between man and the environment. Moreover, it has the potential to create a powerful new style of interaction. In a sense, AR can be viewed as an "Intelligence Amplifier", as proposed by Brooks (Brooks, 1996).

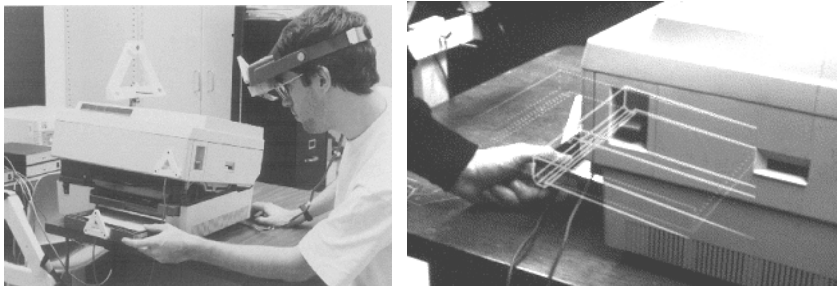
### 3.2 Applications

There are many potential applications of AR, such as medical, manufacturing, maintenance, annotation, navigation, guidance, entertainment, and military applications.

This section presents some example applications in the areas of equipment maintenance, repair, and manufacturing, and nuclear power plant maintenance that are related to this experimental study.

### 3.2.1 Equipment Maintenance, Repair and Manufacture

The maintenance, repair and manufacture of complex equipment is one of the most popular areas of research into applications of AR and several interesting prototype AR systems have been developed in this area. Figure 2 shows an early research prototype AR system for repairing a laser printer (Feiner, 1993). In this project, maintenance procedures were presented to the user step by step as animated wireframe graphics, depending on the progress of the work. The instructions are much more comprehensive than those of the paper-based maintenance manual.



*Figure 2. A prototype laser printer maintenance application in action.*

Figure 3 shows a prototype guidance system to support the assembly of wire harnesses, which are part of the electrical system of an aircraft (Mizell, 2001). While a technician wearing a see-through HMD assembles a wire harness, he/she can see the assembly procedure on the layout board. Because there are thousands of sorts of wire harnesses and each assembly procedure is complicated, an AR-based assembly guidance system is effective for this kind of manufacturing work.



*Figure 3. A prototype wire harness assembly system developed by Boeing.*

### 3.2.2 Nuclear Power Plant Maintenance

Some AR research has been done in the nuclear industry, mainly focussed on maintenance tasks, and many AR researchers view the nuclear industry as one that would benefit most from effective AR systems.

Dutoit et al. of the Technical University in Munich, Germany, have been developing a prototype AR-based maintenance support system called STARS (Sticky Technology for

Augmented Reality System) (Dutoit, 2001). Field operators are equipped with a wearable computer with a HMD. A hand-held LCD panel serves as a secondary display. The HMD is used to augment the operator's surroundings with supplementary information, while the LCD panel is used to display detailed information. The HMD can be flipped up to enable the technician to read the LCD plate, and the LCD plate can be put away when not needed so that it does not get in the way of the technicians work. The content of the presented information is called IETMs (Interactive Electronic Technical Manuals), which is a replacement for conventional paper manuals and checklists. Figure 4 shows an example of the view seen by the operator via the HMD.

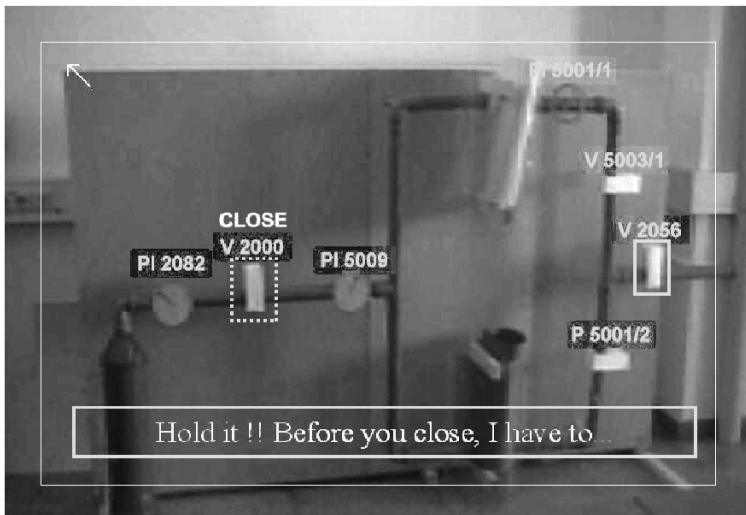


Figure 4. Illustrative (not actual) HMD view of a helium flushing system as seen when using STARS.

Other work in, or directly related to, the nuclear industry includes Branch et al. (Branch, 1999), who have also examined the application of AR technology to radiation protection applications and Pretlove (Pretlove 1998), who have studied the combination of telerobotics and AR for working in hostile environments (primarily for inspection tasks).

Researchers at the Technical University at Clausthal, Germany, have also studied the application of AR to teleoperation for the nuclear industry, as well as for training (Elzer and Behnke, 2001). They have developed an augmented interface that they call a “sticky cursor” that visualises a robot’s degrees of freedom (DOF) and can be used to directly manipulate the robot. The main conclusion of their evaluation of this interface was that visual enhancements through AR compensate for most disadvantages of monoscopic views.

Another project is STARMATE (Schwald, 2001), which is an EU-funded project studying the application of AR technology to computer-guided maintenance of complex mechanical elements. The prototype system under development is intended for use in aircraft maintenance and the maintenance of pumps and turbines in nuclear power plants.

### 3.3 Enabling Technologies

The enabling technologies that form the main building blocks of typical AR systems

(Azuma, 2001) are;

- displays
- tracking
- registration, and
- calibration.

These technologies are used by AR systems utilizing image presentation. When using other modalities, other technologies might be necessary.

### **3.3.1 Displays**

The typical information presentation method for AR systems is visual displays. These can be classified into three categories, (1) HMDs (sometimes called head-worn displays), (2) handheld displays, and (3) projection displays.

The most popular of these displays is the HMD, which is worn on user's head and provides images in front of the user's eyes. There are two main types of HMD: optical see-through displays and video see-through. Optical see-through displays use semi-transparent mirrors to provide a computer generated overlay image over the directly seen background while video see-through displays use video capture from a head-mounted video camera to provide a real background image on which the computer-generated overlay is superimposed. Both display types are usually based on LCD technology because it is small, lightweight and has low power consumption.

Handheld displays are occasionally used in AR systems. For example, Rekimoto has developed a virtual magnifying glass system, which uses a handheld display (Rekimoto, 1997). In the system, a video see-through approach is applied to overlay the augmentation over the real background view. In addition to virtual magnifying glasses, handheld displays are often used as a kind of window to look into the augmented environment.

Projection displays projects virtual information directly onto physical objects. This type of display typically uses a video projector or a laser pointer (Mann, 2000) as a virtual information projection tool. The merit of the projection display is that the user does not normally (in some systems head-mounted projectors are used) need to wear any display equipment. Also, high accuracy of overlay can be easily realized.

### **3.3.2 Tracking**

The term 'tracking' is used to describe both 6 DOF measurements, and 2D-feature correspondence in image sequences. In this section, the term 'tracking' is used to describe 6 DOF measurements.

Tracking a user's position and orientation is important in order to achieve accurate overlay between virtual information and the real world. Several types of tracking technologies have been proposed and are used not only for AR but also VR applications. They are classified into five categories according to their principles (Rolland, 2001).

- Time-frequency measurements  
A radio wave is used to measure the position and orientation of the AR device in the environment. A typical method is GPS (Global Positioning System). The position and orientation on the globe can be measured by receiving radio signals

from several satellites. In the GPS receiver, the time lag between the signals from the different satellites is measured to calculate the position and orientation. A method using ultrasonic signals instead of radio has been also proposed (Foxlin, 1998).

- **Spatial scan**  
A typical example of spatial scanning is to measure the positions of fiducial markers placed in the environment at known locations. A method has been proposed (Welch, 2001) that utilises cameras attached to the user to measure the positions of many LED markers which are located on the ceiling.
- **Inertial sensing**  
Inertial sensing means that the AR device can measure its own position and orientation independently from its surroundings. Typical methods used are accelerometers and gyrocompasses.
- **Mechanical linkage**  
Mechanical linkage methods employ strings or rods to measure position and orientation. They often limit user freedom of movement, however, force feedback can be provided through the mechanical links.
- **Direct-field sensing**  
Direct-field sensing means that the AR system scans the surrounding space and estimates its position and orientation from the scanned information. A typical method is that the system captures surrounding images, extracts features from the images, and recognizes its own position and orientation by comparing them with the environmental model (Neumann, 1999).

The methods listed above are usually combined to improve the tracking accuracy (You, 1999, Yokokohji, 2000). These are called hybrid-tracking techniques.

### **3.3.3 Registration**

Registration is the alignment of computer-generated images with the surrounding real world. In AR, the objects in both virtual and real world must be properly aligned in order to achieve an effective illusion that virtual objects coexist with the real world. However, it is difficult to realise accurate registration, and this problem is one of the technical bottlenecks of AR systems. Registration errors are classified into two types: static errors and dynamic errors.

Static errors occur even when the user's viewpoint and the objects in both the virtual and real world remain completely stationary. The main reasons for static errors are optical distortion, errors in the tracking system, mechanical misalignment and incorrect viewing parameters. Dynamic errors have no effect until the user's viewpoint or a virtual object move. They are mainly the result of system delays or time lags.

### **3.3.4 Calibration**

Vision-based AR systems require adequate calibration to realize accurate registration. Factors to be calibrated may include camera parameters, field of view (FOV) and optical distortion of the HMD, sensor offsets, object locations, etc., depending on the system configuration. Such calibration must be done beforehand, and is often taxing work. A number of methods have been proposed to avoid or reduce the amount of calibration



work required to set up an AR system (Kukulakos, 1998, Seo, 2000, Hoff, 2000).

### **3.4 AR issues**

Except for a few commercial applications, AR is still at a research stage. There are a number of issues that need to be addressed to improve the usability of AR systems and thus enable wider deployment. These are mainly technical issues, but some are also user interface issues and social issues.

#### **3.4.1 Technical Issues**

As described above, AR systems typically comprise of four enabling technologies. However, these technologies are generally not yet mature enough for practical purposes.

- **Displays**  
Concerning image presentation, HMDs should have much wider FOV, be lighter, cheaper, use lower power consumption, have higher resolution, and have better contrast against the outside view in case of see-through displays.
- **Tracking position and orientation**  
The accuracy of the position and orientation tracking is the most serious problem for AR systems, because it causes serious registration errors. In cases where all the information about the surroundings are known or some devices are already prepared in the environment, accurate tracking methods have been proposed. In unprepared environment, however, tracking issues still remain.
- **Registration**  
There are registration issues related to position and orientation tracking. As described above, registration errors cause a serious mismatch between virtual objects and the surrounding real view. Registration errors are caused by a variety of technical factors, not only tracking errors but also distortions in the optical systems of HMDs, time lag during information refresh rate, etc.
- **Calibration**  
Most AR systems require calibration before using. Although it is necessary in order to improve the accuracy of registration, calibration forces the user to do perform a typically complex and challenging operation. Calibration free or auto-calibration approaches are desirable.
- **Other issues**  
Wearable AR systems have a number of issues, in common with wearable computers in general. These include weight, power consumption, system performance, comfort, not impeding actual work, and so forth.

#### **3.4.2 User Interface Issues**

By combining generated virtual objects with the real world, AR systems achieve an illusion that is not a natural environment for humans. User interface issues are therefore important to attain easy perception and natural operation. Most research on this issue has dealt with low-level perception problems, such as how accurate registration should be, or how short the refresh time lag should be. However, one of the main uses of AR systems is to support complex tasks through information presentation. This means that the AR system should provide high-level information to the users in a manner that

ensures that the system is easy to use. Although some recent research into AR systems has dealt with usability issues (Rauterberg, 1996), many more studies are required because a variety of different types of AR system are being proposed and the usability of these systems depends on each system configuration and on the target tasks of the users.

### **3.4.3 Social Issues**

When a new technology spreads over the society, it should be acceptable to the people. Although this issue is important, there is little research concerning social acceptance. The issues of social acceptance depend on both the AR system itself and the situation in which it is used. For example, in the case of an AR system that is to become an integral part of the user's daily life, it is important that the system has a high level of usability, is reasonably cheap to buy and maintain, ensures personal privacy, and even has a fashionable appearance. On the other hand, in case of systems for professional use, operability, task workload, efficiency and heavy-duty design are important factors.

## **4. HUMAN PERCEPTION IN MULTIPLE MODALITIES**

The limitations in human attention are critical for information processing. A number of psychological and physiological theories have been proposed to explain the mechanisms underlying attention, most of which involve the concepts of limited resources and selection.

The question of concern in the study presented in this report is how attention should be directed to maintain optimal control of the situation surrounding an operator in a hazardous environment, and how this should be presented to achieve maximum focus on the important information. More specifically, what cues or combination of cues are best suited to giving a field operator effective comprehension of the surrounding radioactivity without interfering with primary work tasks.

### **4.1 AR and perception**

Nature has provided human beings with five senses. However, the computer technology has generally restricted its users to using their sense of vision, and, more recently, receiving auditory access to information. The other three senses - touch and chemical (both taste and smell) have not been put to much use in computer-based applications. Common for all modalities are that they can be used to direct our attention in a specific direction.

Given the anticipated developments in AR computer technology in a five to ten years perspective, visual, auditory, and haptic presentation are the most appropriate to consider for the presentation of radiation level information in a manner that optimally supports human behaviour. Defining "optimal" as behaviour that will maximise expected value and minimise expected cost (Wickens and Hollands, 2000), the choice should be made on basis of what best supports the operator's radiation awareness (RA) with minimal interference with the operator's work in the NPP. I.e., how the operator should build a mental model of the radiation without having to spend too many cognitive resources focusing on the radiation at the expense of other tasks.

A critical factor for building this mental model is how the spatial layout is presented and

supports the human perceptual system.

## 4.2 Visual superiority

The visuospatial code receives inputs from the tactile organs in the skin as well as from the eyes. Accordingly, the shape representation of each tactile input could be compared with the shape representation of the visual input. The visual and tactile sensory modalities are calibrated so that the same representation of the shape of an object is constructed on the basis of either visual or tactile information. Since this calibration is very good in people with no tactile or visual deficits, a task like tactile recognition of objects is easy (Glass and Holyoak, 1980). Vision is generally considered the dominant sense, and there is evidence that human cognition of spatial location is oriented around vision, with people often using visual imagery as mediating representations of thought (Rock and Victor 1964, Kosslyn, 1980, 1994). The strong or complete dominance of vision over touch has been demonstrated in a number of varied perceptual tasks, including judgement of size, curvature, length, depth, texture, and spatial location (Glass and Holyoak, 1980). However, the degree of visual dominance is believed to be dependent on the task. It is more likely that vision will dominate if conflicting information is of a kind that is normally obtained visually.

Visual and auditory spatial information is believed to have the same relation as visual and haptic information. In the context of spatial conflicts, vision strongly dominates audition and the other spatial senses under a wide range of conditions (Walker and Scott, 1981). However, for duration and amplitude the visual dominance over auditory information is not as clear as the one over haptic information. Presenting tones and light signals to the subjects, Walker and Scott (1981) found that the intensity of the signals was determining what modality would be dominating in a reaction time task. If the light signals were presented with high intensity and the sound signals with low intensity, then vision seemed to be dominating, and vice versa. Synchronization is also believed to influence the visual dominance. Bertelson and Aschersleben (1998) found that desynchronising the input eliminated the attraction to visual targets when presenting conflicting visual and auditory information.

According to Warren (1979) and Pick et al. (1969, both cited in Walker and Scott, 1981) the haptic modality dominates the auditory for spatial perception. Further support for this view is found in Kerzel (2001), which argues for an interdependency of visual memory and haptic perception, while acoustic signals does not influence the visual memory.

One argument for focusing on visual information compared to auditory and haptic could be the impression of continuity. It is easier to present a full overview of a larger area using a map, compared to sound and vibrations. A finding supporting the view that a full overview is important is the development of an acoustic navigation system for blind people (Shoval, Borenstein and Koren, 1998). Blind or visually impaired people testing the system preferred the mode that gave the most complete overview of the situation, except when they were located in well-known surroundings.

Emphasis should be placed on consistency between different modalities in the presentation of spatial radiation information. Two classic experiments performed by Brooks (1968) and Baddeley and Lieberman (reported in Baddeley, 1976) suggest that if

different spatial input is presented from two different modalities, a spatial conflict will be present. The important finding in these experiments underline the need for careful presentation of the same information, and that conflicting presentation could lead to inhibition of what one really want to be learned.

### **4.3 Research issues**

Using the AR system prototype developed for the experiment described in this report, information about radiation is presented to the subjects using different modalities. On the basis of the presented background in the technology review, human attention limitations and results from research on different modes for perception, the visual mode was chosen as the main mode and compared to a combination of visual, auditory and haptic presentations. The goal of the study was to examine the effect of the different presentations and consider how this will have impact on future decisions on the development of multi-sensory presentation technology. Presented with sufficient consistency, the multi-modal presentation should lead to better understanding of the radiation distribution. On the other hand, multi-modal presentation could lead to greater demands on attention and thus degrade performance in situations where conflicting information is presented.

The main purpose of the second part of this study was to evaluate the usability of the prototype. In addition to the issue of presentation, the use of the HMD was tested for on site immediate collection of human performance data, compared to a delayed condition. The purpose of this test was to examine the feasibility of computer mediated testing in field studies.

## **5. METHOD**

### **5.1 Participants**

Eighteen subjects (17 male, 1 female) from the Man, Technology and Organisation (MTO) and Reactor Operation and Engineering (ROE) divisions at the HRP participated in the experiment. Subjects' age varied from 29 to 47, with an average of 36 years. Six of the subjects were not familiar with the reactor hall, while twelve work there daily with an average of 8 years experience. 5 of the subjects used glasses during the tests, while 1 used contact lenses. Two of the subjects reported colour blindness, and no subjects reported problems with the auditory or visual capacity.

#### **About operator doses**

The extra doses subjects received during their short stay in the reactor hall were below what can be measured, and have been calculated to approximately 10  $\mu$ Sv. The study was performed in accordance with the HRP quality assurance procedures for studies involving human participants and approved by the HRP Committee for Human Research Subjects.

### **5.2 Experimental design**

The experiment had two scenarios, consisting of one task each. The first scenario was

designed to test the modes of radiation presentation. In this scenario, a between subjects repeated measures factorial design was used, with presentation of radiation as the independent variable.

The independent variable in scenario 1 had three levels. Each of three groups of six subjects had radiation presented in different ways according to Table 1 below. The two AR groups were presented with sound and vibrations for either the first (ARSVF) or last (ARSVL) part of scenario 1. The MAP group used a paper map of the reactor hall for both parts of scenario 1 (maps are presented in Appendix 3).

Table 1. Experimental design.

Condition	Measurement task	B	Measurement task continued
MAP	Map presentation	R	Map presentation
ARSVF	HMD + sound + vibrations	E	HMD
ARSVL	HMD	A K	HMD + sound + vibrations

The second scenario was a scenario for testing the measurement of radiation awareness and usability of the equipment. The radiation awareness test had a balanced presentation of test material with two levels (immediate and delayed) on the repeated factor. In this scenario, all subjects performed the same test without any manipulation of tasks or equipment.

### 5.3 Dependent variables

Several performance measures were collected to reveal potential differences between conditions. Data was collected using questionnaires and by measuring operator positions relative to the radiation sources. Additional comments were also collected from the subjects to support the interpretation of the results.

#### 5.3.1 Radiation Awareness

A score of the subjects' awareness of the radiation profile of individual locations was obtained. The scoring technique has previously been tested in VR-based training (Nystad, Drøivoldsmo and Sebok, 2002) and will be documented in Drøivoldsmo, Sebok and Nystad (2002). Assessment was done by presenting a set of photographs of the reactor hall to the subjects, like the example provided in Appendix 1 and asking the subjects where the highest or lowest levels of radiation would be found. For each test, the subjects were shown a sequence of four pictures and asked to indicate which of three positions (i.e., A, B, or C) was associated with the highest or lowest radiation level. The answers were scored as either right or wrong, and RA was then calculated as the number of correct answers for that sampling period. Thus, RA scores varied between 0 and 4 with higher scores indicating a higher degree of awareness of the radiation field.

For both scenarios, the correct answers were scored using the program VRdose (Iguchi and Rindahl, 2002) by extracting the exact measures in  $\mu\text{Sv}$  and finding the highest or lowest radiation point. For scenario 1, the values obtained using VRdose were compared to the paper map (direct measurements in the reactor hall) and no differences were found.

### **5.3.2 Route deviation**

For the second scenario a measure of the best path was scored. An optimal path was obtained using VRdose. A VR model of the reactor hall was loaded into VRdose and a manikin was placed in the virtual environment at the six checkpoints the subjects had to pass on their route out of the reactor hall. Each subject's deviation from the passing point was then scored. The distance between the checkpoints was equally distributed along the total length of the route each operator had to walk out of the contaminated area. The deviation from the optimal path was obtained from video recordings of each subject's movements, transferred into a map, and measured, using the V&V VR measurement tool (Drøivoldsmo et al., 2001).

The effect of error direction was approximately equal, and a root mean square (RMS) score (Buck, Zellers and Opar, 2000) was used for scoring the correctness of the choice of escape route out of the reactor hall. The first two and the last of the six passing points had a variation in radiation levels of less than 4  $\mu\text{Sv}$ . and it was judged as highly unlikely that the subjects would be able to detect variation as low as this from the resolution of the information presented. These three checkpoints were therefore excluded from the analysis.

In addition, a measure of obtained dose was calculated, under the assumption that subjects maintained equal speed out of the reactor hall.

### **5.3.3 Subjective situation awareness (SA)**

A three dimensions version of the Situational Awareness Rating Technique (SART) was used as basis for assessment of subjective situation awareness (Gawron, 2000; Taylor and Selcon, 1991). In addition a fourth question, asking subjects to directly rate their own understanding of the situation was administered (Gravelle and Schopper, 1993). A five point Likert-scale was used and inverting questions 1 and 2, and using the average of the three dimensions SART calculated the score. Thus, SA scores varied between 0 and 4 with higher scores indicating a higher degree of self rated awareness, learning the radiation distribution in the test scenarios. Printout of the questionnaire is provided in Appendix 2.

### **5.3.4 NASA-TLX**

Operator workload was measured after both scenarios, using the NASA Task Load Index (Hart and Staveland, 1988; NASA, 1988). The raw scoring technique discussed in Byers, Bittner and Hill (1989) was used for collection of the workload data.

### **5.3.5 Usability**

A 7-item questionnaire was developed to evaluate system usability. Four items were taken from Brooks' (1998) usability test and one item was an assessment of the usefulness of the AR system. All of these five questions used a five-point Likert scale, and the subject was asked to indicate the amount of agreement with these statements, from "Completely agree" to "Completely disagree". The last two questions were open-ended and asked what the subject liked about the system and what should be improved.

## 5.4 Experimental setup

### 5.4.1 Materials

The reactor hall is located inside a mountain and is accessed via a 40 metres long tunnel. The main hall, which is level with the external entrance and the tunnel, measures approximately 28 m by 10 m and is 10 m high. The subjects were tested in the main hall. The tracking equipment and experimental control station were placed on a gallery 4 m above the main hall floor with a free view of the tracking area in which the subjects performed their tasks.

### 5.4.2 Technical setup

For a detailed description and discussion of the technical setup the reader is referred to Reigstad and Drøivoldsmo (manuscript in preparation). The description in this report covers only the details that are necessary to understand how the experimental data was acquired.

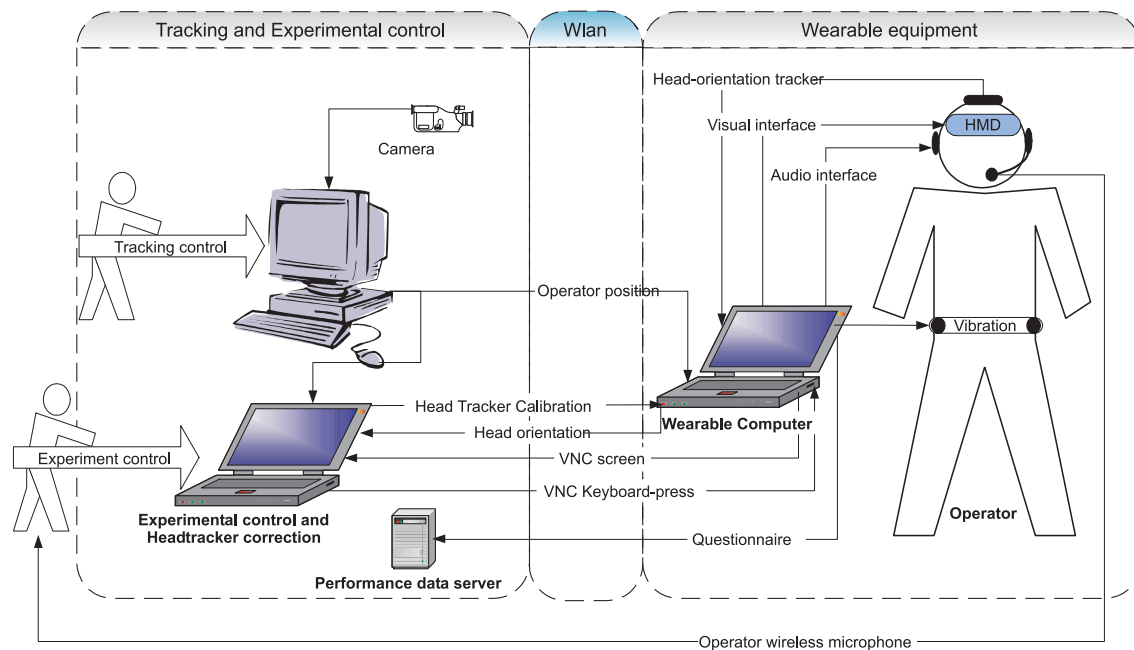


Figure 5. The system used three computers for the AR system and an additional server computer to safely store the human performance data collected. A wireless local area network was used to transfer data between the wearable equipment and the other computers. Additionally, the operator had a wireless microphone for speech recognition during questionnaire data collection.

### Position tracking and Experimental control

The left side of Figure 5 illustrates the stationary equipment used during the experiment. One computer (“tracking control” in the figure) had a video camera attached and was used to support tracking of the operator’s position in the reactor hall. A second, laptop, computer (“experimental control” in the figure) was used to control the wearable equipment on the operator.

## Wearable equipment

The right side of Figure 5 illustrates the wearable equipment used by the subjects during the experiment. This equipment was worn as shown in Figure 6.



*Figure 6. The figure shows the wearable equipment consisting of a laptop in a backpack and a HMD with integrated audio. The head-orientation tracker is mounted on top of the HMD's display unit.*

A powerful laptop computer was used as the wearable computer by placing it in a backpack. Communication between this wearable computer and the other computers in the system was achieved using a wireless network.

A HMD served as the wearable computer's display. As the HMD was semitransparent, it was possible to see the computer screen image superimposed over the subject's view of the real world, however the subjects thus viewed the reactor hall as if wearing dark sunglasses. The HMD displayed either the AR radiation map representation or the questionnaire system, depending on actions performed by the user of the experimental control computer, which is described below. Audio generated by the software running on the wearable computer was played through stereo earphones integrated in the HMD.

A head-orientation tracker was mounted on HMD and connected to the laptop. When running the AR radiation map application, the head-orientation tracker was used to correctly display the radiation visualisation depending on the orientation of the subject's head.

The wearable computer also controlled two vibrators mounted on a belt around the subject's waist. With the belt on, the subject had one vibrator on each side of his body, as illustrated in Figure 5.

## Tracking control

The tracking control computer displayed a video image on its screen that provided an overview of the entire experimental area within the reactor hall. This was used to



determine the location of the subject and to transmit the calculated position to the wearable computer and the experimental control computer via a wireless network. The process of determining the exact position of the subject required human supervision. The user of the tracking control computer used the computer's mouse to track the image of the subject in the video with the cursor. The computer used the position of the cursor over the video image to calculate the subject's location in the reactor hall. The wearable computer was thus continually updated with real-time position data, enabling it to always display information to the subject that was correct in relation to the subject's actual position in the reactor hall. Note that since this tracking system required human intervention and is also limited by the resolution of the video image and computer screen, it is neither the most accurate nor the most efficient solution (see the comments on this in the Results section of this report) but was considered sufficient for the purpose of the experiment.

## **Experimental control**

The experimental control computer was used to run two applications that were required to control the wearable computer via the wireless network

The first application displayed a VR model of the reactor hall. This application was active only while a scenario was running. The application displayed a view of the VR model that was continuously updated with the operator's current location from the tracking computer and the operator's head-orientation from the head tracker, so that the application displayed a VR representation of the subject's view of the reactor hall. The user could monitor the view of the VR model and visually determine whether the head orientation tracking system needed recalibration. When necessary, the keyboard on the experimental control computer was used to manually recalibrate of the head-orientation tracker.

The second application on the experimental control computer was a human—mediated speech recognition system that was used to control the questionnaire system. With this application, it was possible to take complete control of the wearable computer (“VNC key press” in the overview Figure 5) and view the exact same screen image that the operator (“VNC screen” in the overview Figure 5) was seeing in the HMD. Wearing a headset to hear the subject's verbal responses and commands, the user of the experimental control computer was able to fill in the questionnaire on behalf of the subject.

## **5.5 Scenarios and procedure**

### **5.5.1 Scenario 1**

In the first scenario, the operators were supplied with a map of the reactor hall indicating six locations where they were to find an analogue meter and record the values displayed on it. On the map, a numerical sequence from 1 to 6 indicated the order that the subjects should use for the registration. In addition to the first map, the operators were also supplied with a sheet of paper with photographs of the exact locations of the meters. The radiation maps used in this scenario were based on a detailed measurement of level one in the reactor hall done prior to the experiment. It is presented in Appendix 3.

The first test of RA was performed before the subjects were given any information about the radiation levels in the reactor hall (Table 2). This test was done outside the reactor hall. After a training session lasting approximately 10 minutes, scenario 1 started and the three first registrations of readings were performed. Then there was a break in the scenario and RA and SA data were collected. For the AR groups, the tests were presented on the HMD and filled in with the assistance of the user of the experimental control computer, using the technique described above. The MAP group used the experimental control computer directly to fill in the questionnaires. After finishing the last part of the scenario, data was collected in the same way as for the break. During all data collection in the reactor hall, the information about radiation levels was removed, but subjects were allowed to examine their surroundings.

Table 2. Data collection in the first scenario

Pre experimental		Experimental			
RA1 (pretest)	Training in AR or MAP conditions	Measurement task	RA2  SA1	Measurement task continued	RA3  SA2  TLX1

### 5.5.2 Scenario 2

In the second scenario, the operators were presented with a new situation where fictionally high levels had been added at three locations in the radiation map. 1) In the fuel rod gas blow system, control valve VA2038. 2) In the ion exchanger F172. 3) Besides the rotating disk, between Rack A and Rack B (the radiation map is presented in Appendix 3).

All operators used the AR radiation visualisation of the new radiation distribution, and were given approximately 10 minutes to familiarise themselves with the new radiation map. They were further instructed to walk to the far end of the reactor hall and use the AR visualisation to determine a route out of the reactor hall that would minimise their exposure to the fictional radiation.

Table 3. Data collection in the second scenario. The navigation task is the task of walking the optimal route out of the reactor hall.

	Immediate	Delayed
Navigation task	RA4	RA5
	SA3	SA4
	TLX2	Usability

After walking from one side to the other of the reactor hall, subjects were asked to fill in the RA, SA and Workload questionnaires (Table 3) the same way as for the AR condition in scenario 1 (immediate condition). After a time delay of 10-15 minutes (the time it took to be cleared out of the contaminated area), the operators answered a new set of tests on a desktop computer in an office environment (delayed condition).

## 6. RESULTS

### 6.1 Data handling

Scoring of radiation awareness and input of data to the statistical analysis programme was done automatically from the Halden Questionnaire data collection engine (Drøivoldsmo and Johnsen, 2002).

Analysis of variance was performed to identify differences among the different forms of presentation for RA, subjective SA. All analyses were performed in the Statsoft, Statistica version 5.5. Due to the explorative nature of this experiment and to prevent type II errors, post hoc testing used the LSD test, which is comparable to using a t-test for independent samples without using Bonferroni correction. There was one missing data point for the second scenario route deviation score for which mean replacement was used.

No differences in scores or interactions were found for any of the measures due to the background (high or low frequency of work in the reactor hall) variable. However, test of the RA pre-test showed border significance ( $F(1,16)=4.3$ ,  $p=.055$ ) in favour of the experienced group. This variable has therefore not been implemented in the analysis.

## 6.2 SCENARIO 1

### 6.2.1 Radiation awareness

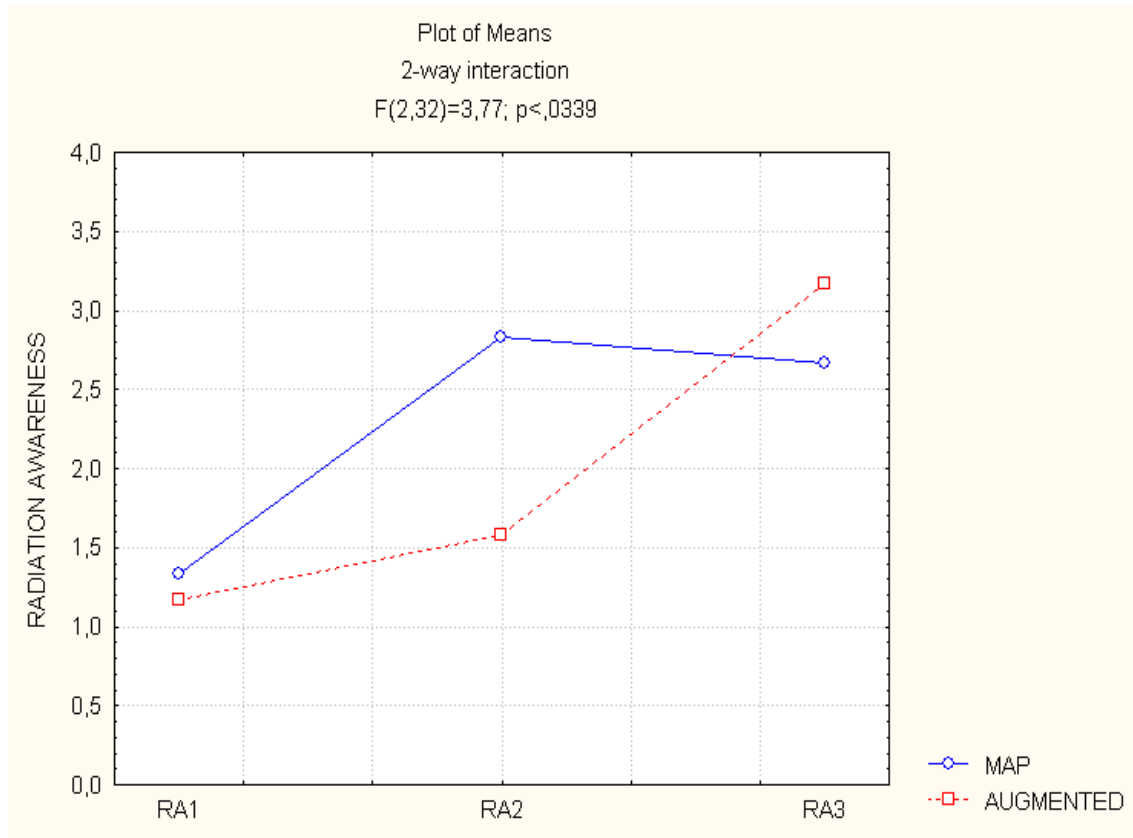


Figure 7. Interaction between the map and augmented conditions for repeated measurements in scenario 1.

The RA data was tested in a between groups (map or augmented presentation, 2 levels) repeated measures (time for test, 3 levels) analysis of variance. The overall analysis showed no main effect from the independent variable presentation. However, there was an interaction between presentation and repetition ( $F(2,32) = 3.77, p < .05$ ) indicating a different learning curve in the two conditions (Figure 7).

#### Comparing map, sound and vibrations first, sound and vibrations last

Splitting the augmented level (Figure 8) of the factor presentation into ARSVF and ARSVL (see table 1) indicated that the ARSVF group had a different learning curve than the ARSVL group ( $F(4,30) = 4.32, p < .01$ ). Post hoc testing of the second RA (RA2) measurement showed a significant difference both between the MAP group and the ARSVL ( $p < .01$ ), and the ARSVF and the ARSVL group ( $p < .05$ ).

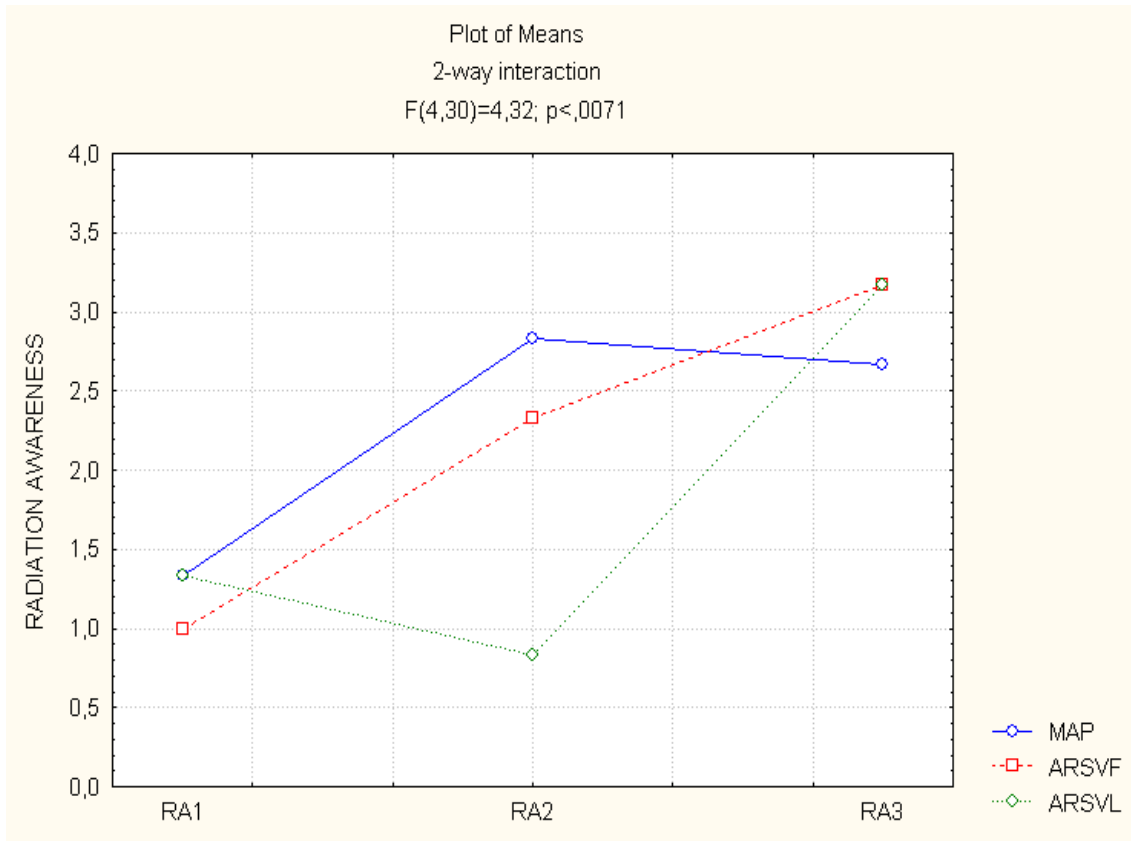


Figure 8. Interaction between independent variable presentation and repeated testing

### Analysing the augmented condition

Analysis of the augmented group showed an interaction between the ARSVF and ARSVL group,  $F(2,20) = 3.53, p < .05$ , indicating that presenting sound and vibrations for the first or last part of the scenario gave different RA-results for the two conditions.

There was a learning effect of repetition ( $F(2,30) = 20.64, p < .001$ ), showing that all groups learned the localisation of radiation better through working with the material. Means and standard deviations are presented in Table 4.

Table 4. Means and standard deviations (in parenthesis) for the RA tests

Group	RA 1 (pre-test)	RA 2 (scenario break)	RA 3 (after scenario)
Map	1.33 (1.03)	2.83 (1.17)	2.67 (1.03)
ARSVL	1.33 (1.03)	.83 (.75)	3.17 (.98)
ARSVF	1.00 (.63)	2.33 (1.21)	3.17 (1.17)
All groups	1.22 (.87)	2.00 (1.33)	3.00 (1.03)

### 6.2.2 Subjective Situation Awareness Ratings and workload

A between groups (map, augmented) repeated measures (SA1, 2) analysis of variance for the SA questionnaire data did not reveal any significant differences. Splitting the augmented group the same way as for the RA data showed a significant interaction ( $F(2,15) = 4.17, p < .05$ ) for presentation and repetition (Figure 9).

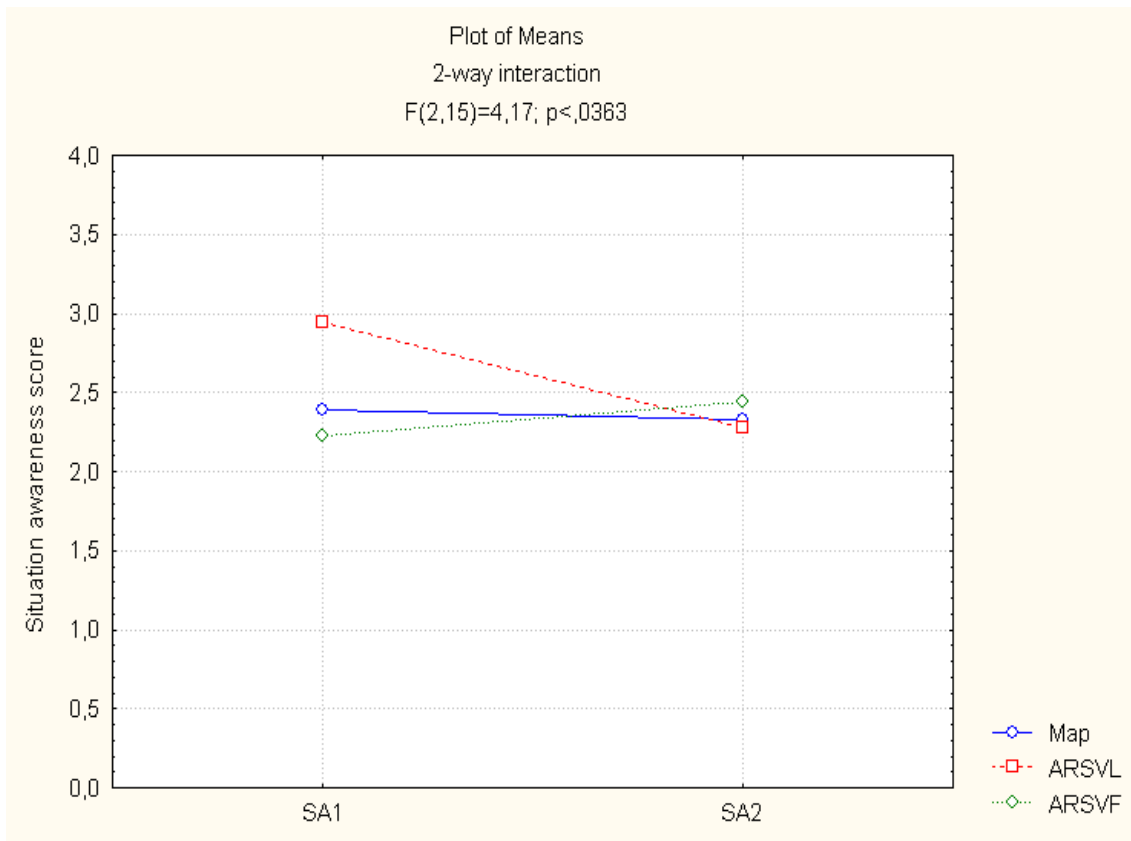


Figure 9. Subjective ratings of SA tested after task one and task two respectively.

Post hoc testing showed that the ARSVL group differed from the other two groups for the first test of subjective SA (SA1).

The NASA-TLX workload measure did not show any differences between the different modes of presentation. However, a correlation was found between NASA-TLX and the subjective SA measure collected at the end of scenario 1 ( $r = -.59$ ,  $n = 18$ ,  $p < .01$ ). This correlation is explained by the demand factor of SA and confirms the findings from Selcon, Taylor and Koritsas (1991).

### 6.3 SCENARIO 2

In this scenario all subjects performed the same task as a basis for the usability ratings and investigation of the test methodology.

No effect of experience was found between the immediate and the delayed RA-test. There was no interaction between the balanced tests, indicating a stable measurement instrument, independent of test order, but there was an overall effect of forgetting from the immediate to the delayed condition ( $F(1,16) = 4.77$ ;  $p < .05$ ) where performance in the immediate condition was better than in the delayed condition. A simple effect contrast analysis for the immediate test confirms what can be seen in Figure 10, and shows a difference between the augmented and the map groups ( $F(1,16)=4.94$ ;  $p < .05$ ). At the same time the augmented group reported higher workload ( $M=3.9$ ) than the map group ( $M=2.5$ ), ( $F(1,16)=8.85$ ;  $p < .01$ ). These results is in coherence with the results

found in scenario 1, where a learning effect could be seen, and the fact that the augmented group was more familiar with the AR system could have made them more capable of putting effort into using the information provided by the system. The SA score did not differ from the immediate to the delayed condition.

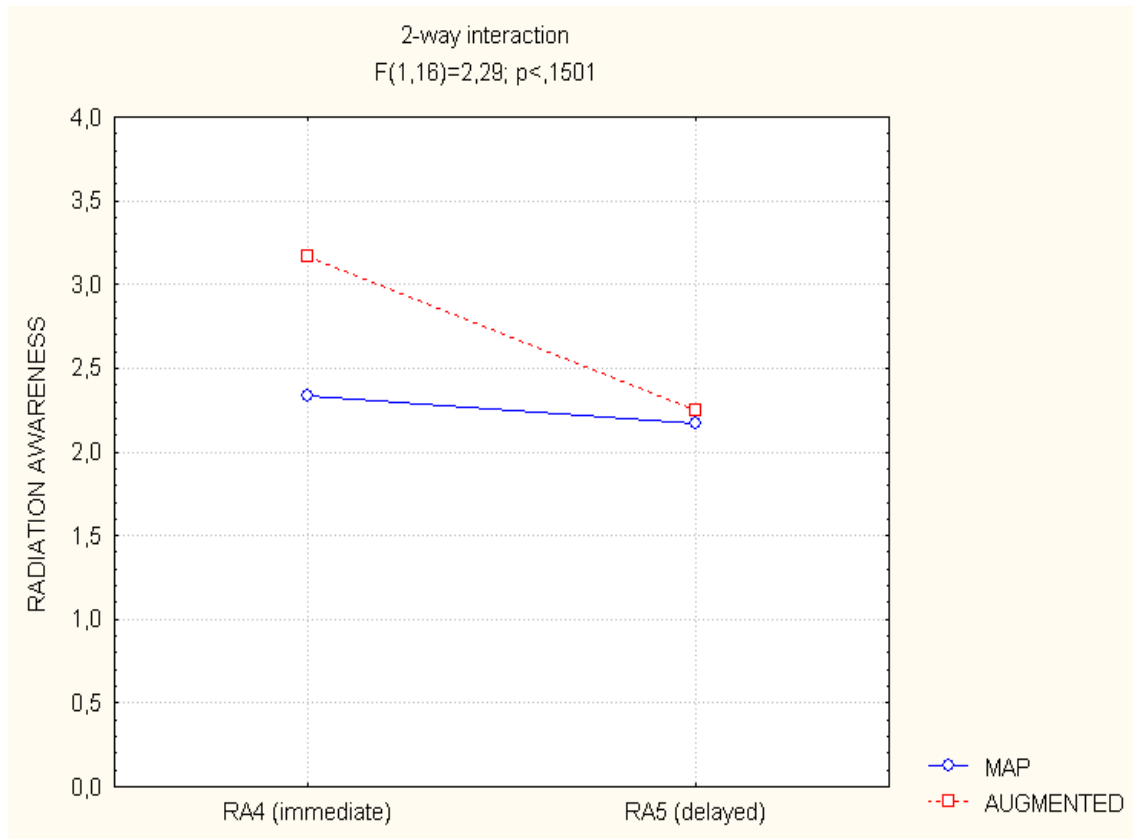


Figure 10. Immediate and delayed testing of RA.

The results concerning immediate vs. delayed testing will not be discussed further in this report.

The RMS optimal route deviation score and the total dose score did not show sensitive to the independent variable from scenario 1. The means and standard deviations are reported in Table 5.

Table 5. Means and standard deviations for the deviation score. Higher values indicate higher deviation from the optimal route. Measurement units are in centimetres. Mean replacement has been done for one of the cases in the ARSVL group.

	Mean (SD)	N
MAP	96.0 (47.7)	6
ARSVL	93.6 (35.4)	6
ARSVF	81.5 (50.4)	6

## 6.4 Usability statistics

After finishing the practical test, a five question Likert-scale usability questionnaire was



presented to the subjects (Appendix 4). An overall comparison of subjects with experience from working in the reactor hall showed no difference between the two groups for the first four questions. The final Likert-scale question, which was about the “usefulness of such a system in my daily work”, was relevant only for the group of experienced operators. Of the 12 operators that answered this question, 11 agreed or fully agreed that the system would be very useful. Descriptive results from the questionnaire are shown in Table 6 below.

*Table 6. Means and standard deviations for the usability questions. Scores ranged from 1 (negative) to 5 (positive). For questions 1 and 3, scores have been reversed.*

Questions	Mean (SD)	N
1. I needed to learn a lot of things to get started with the system	3.5 (1.1)	18
2. I found the various functions in this system were well integrated	3.1 (0.7)	18
3. I found the system unnecessarily complex	3.9 (0.5)	18
4. I felt sure in how to use the system	3.1 (1.0)	18
5. I think this system will be useful in my daily work	4.3 (0.6)	12

## 6.5 Qualitative usability data

All subjects were asked to write, in their own words, what they found to be the strong and weak sides of the system. The qualitative data set was easy to interpret, surprisingly consistent, and supplied valuable advice on what to improve in future developments of the system. Comments could be interpreted in the following five categories:

- Visual radiation presentation
- Implementation of the augmented presentation
- Sound and vibrations
- Hardware and system performance
- Ergonomic issues

### 6.5.1 Visual radiation presentation

The most striking observation from the data was that all 18 subjects commented that the graphical visualisation of the radiation levels was a good way of presenting the information. In light of the open knowledge of the purpose of the study, this could be interpreted as a Rosenthal-effect, but given the high level of practical experience the subjects possessed, such an interpretation would be unlikely. A moderate and more plausible interpretation would be that the visualisation used in this study is better than current techniques for representation of radiation levels at the HBWR. This could be caused by inadequacy in current techniques or that the subjects actually perceived the system as a good system. Both these explanations provide a good foundation for the

visual representation in further developments of the system. Another observation supporting these interpretations is the fact that experienced workers often prefer the conventional method when new technology is introduced, unless the new system shows clear benefits.

The “visual” category was also the category with the most ideas for improvements. Using the 2D landscape metaphor as implemented to present radiation levels produces an undesired effect of hiding information. In the same way that a high mountain obstructs the sight of lower mountains behind it, areas with lower radiation are overshadowed by high peaks in the visual display. Using a transparent wire-framed radiation landscape was not giving the right impression of “peaks hidden behind other peaks”.

Also for situations without this cover-up/shadow issue, subjects reported that interpretation of depth in the radiation landscape was not easy. The problem appears to be that it is difficult to judge the relationship between the real environment and the augmented display when peaks are located at some distance from the subject. The most plausible explanation of this observation is lack of calibration (position or rotation) with reinforced error on bigger distances.

### **6.5.2 Implementation of the augmented presentation**

The Sony Glasstron HMD restricts the view for the operator. This HMD has a 1-2 cm wide rim, and this obviously restricts the operator’s peripheral view of the real world environment, forcing the operator to turn his head more frequently than in a situation with unrestricted vision.

For the radiation map, the geometric FOV was set to 50 degrees. This is a relatively narrow FOV, and according to Waller (1999) subjects would probably be more comfortable with a FOV approximating 70-80 degrees. The restriction in FOV of the radiation map is a result of the narrow FOV of the HMD that it needs to match to provide as accurate as possible AR registration, thus a HMD with wider FOV would alleviate this problem.

### **6.5.3 Sound and vibrations**

Out of the 18 subjects 10 commented on sound and vibrations. A number of 4 positive and 6 negative comments were given. Common for all positive comments was that they saw sound and vibrations as an extension to the visual representation, not just as a stand-alone source of information for navigation. Two subjects reported that the vibrations were directly responsible for them discovering a radiation source that they would not otherwise have been aware of.

The negative comments were concerned with sound quality, volume levels, type of sound signal, and problems of interpreting which radiation sources produced higher or lower levels of sound. Further discussion of these views can be found in the general discussion section.

Based on the usability data, future use of sound and vibrations should be focused on providing warning signals and detection of high radiation levels as a support function for the visual presentation.

#### **6.5.4 Hardware and system performance**

Updating the AR model with the operator's position in the reactor hall was done using a semi-automatic position tracking system. Eight subjects reported that they experienced a delay in the update of picture, sound and vibrations, at some point during the experiment. This is a known problem with the current system and will probably be eliminated when a more accurate head-tracker and a operator positioning system has been developed

#### **6.5.5 Ergonomic issues**

In addition to the previously mentioned FOV restrictions caused by the Glasstron HMD, 4 of the 18 subjects commented on the weight of the equipment. The total weight of the system was 6.1 kg, and it was found usable for the purpose of the experiment. For technical, economical, and practical reasons, the issue of weight was not prioritised in the development of the prototype system. For potential industrial implementations of such systems in the future, the weight should probably not exceed 3-400 grams.

### **6.6 Insights from development of the technical solution**

The AR software system was mainly developed using the Java programming language using the Java 3D framework for the presentation of graphics. Video capture for the tracking system was developed using the Java Media Framework. The interface to the head tracker was written in the C programming language and a Java interface to this C code was developed in order to integrate it with the rest of the system. In a project like this, where programming and development of different software modules is done at different locations and by two different organisations, using free software has been a major advantage, that avoid licensing cost and thus any related version compatibility issues.

However, the operating system used on the computers in the experiment was a proprietary system, Microsoft Windows 2000. Hindsight reveals that another operating system would have been better suited for communication between the computers and may have reduced the issues that were experience with the wide-area tracking (video hardware compatibility issues that required extra development effort to find workaround solutions) and correction of drift in the head tracker.

## **7. DISCUSSION**

Presenting sound and vibration in the first part of the task, scenario 1, facilitated learning in the augmented condition. This finding is consistent with the reports from the usability test after scenario 2, and the absence of negative effects from the sound and vibrations, supports a conclusion about no interference between the different sensing modalities.

A common distinction used when describing human vision is between focal vision and ambient vision. Focal vision uses the centre of the visual field and is able to detect details, while ambient vision makes use of the entire visual field, especially the periphery (Gaulin and McBurney, 2001). Since ambient vision is used for detection of important information, but the focal vision is needed to obtain details, we need to move

the line of gaze to understand what is going on in the periphery of the FOV. Using a HMD restricted to a FOV of only 50 degrees removes much of the peripheral vision and the awareness of peripheral radiation peaks presented to the ARSVL (Sound Vibrations Last) group could have been reduced due to lack of the cues provided by sound and vibrations for the first part of scenario 1.

Looking at the rapid recovery of the ARSVL group getting the sound and vibrations in the last part of the first scenario indicates that some kind of learning took place even during the first phase of the scenario. Automatic processes encoding the spatial relations in the radiation landscape could contribute to an explanation of this effect. Automatic processes do not require attention and are unavailable to consciousness (Shiffrin & Schneider, 1977). However, automatic encoding of spatial locations takes place continuously and when eventually reaching consciousness the information can be recognised.

While performing significantly worse on the RA test, the ARSVL group reported higher SA for the first part of the first scenario. This finding is quite logically explained by the fact that as they had not seen the complexity of the situation, they perceived it as simple, and thus rated their own SA highly. In any self-rating instrument, unattended, forgotten and inaccessible information will avoid capture (Taylor and Selcon, 1991). It should also be noted that the underlying questions used for the SA measure are not constructed to capture situation specific knowledge about what is known. However, the understanding-dimension of the measure is believed to correlate with knowledge scores. Despite the low n in this study, the understanding-dimension had a border significant negative correlation with the RA knowledge measure, strengthening the interpretation that the ARSVL group did not perceive the complexity of the situation.

## **7.1 Accuracy of the AR representation**

In general it has been found that human pattern perception and recognition is quite sensitive to the orientation of the stimuli. However, a number of studies have only found large effects of orientation during the early parts of experiments, indicating that subjects learn something over the course of an experiment and that that enables them to become increasingly insensitive to variations of stimulus orientation (Jolicoeur, 1992). Looking at the operators' complaints, and considering that initial testing exposed inaccurate updates of the model's orientation, this could account for some of the slower learning in the augmented compared to the map condition. If AR registration had been perfect then one would expect operators to acquire more accurate mental representations of the environment more quickly. However, if the radiation levels had not been static or if operators had not had enough time to become familiar with the radiation environment, and thus operators would not have been able to learn to compensate for inaccuracies in registration over time, then poor accuracy could potentially result in inappropriate judgements with serious consequences. The accuracy of the AR system is therefore an important issue that must be addressed in future before such systems can be deployed in hazardous situations.

## **7.2 General discussion**

Divided attention and parallel processing is often considered to be good behaviour in

situations demanding human supervisory control. Operators in complex control environments such as aviation, air traffic control, and process industry need to have the capacity to gather information from multiple data sources. The problem often arises when all sources requiring divided attention are made mandatory rather than optional (Wickens and Holland, 2000). In this experiment it was a concern whether the implementation of three different modalities for presenting radiation would be conflicting and thereby disrupt the perception of the operator. This does not seem to be the case, but the low fidelity of the dependent measures does not permit any conclusion. An evolutionary psychologist would probably state that this is the way nature has solved the problem of conflicting information from different senses. Visual dominance will suppress any conflicting information, even when visual information received is wrong.

### **7.3 Conclusion**

The prototype system implementation and experiment were successful in providing valuable information about technical limitations and detecting the level of accuracy needed for systems used to process control activities. The prototype was found sufficient for the purpose of testing operator radiation awareness.

Based on the results and experiences from this study, priority should be given to the development of the HMD presentation of radiation spatial distribution and amplitude. Improvement of the graphics must be done in conjunction with further testing of FOV and egocentric vs. exocentric view. Sound and vibrations seems to play an important role for how fast the operators learned the radiation layout in the test environment, but mainly as an alarm function for directing the operators' attention.

To avoid the uncertainty about accuracy seen in scenario two, the wide-area tracking system used for the experiment should be replaced with a more accurate fully automatic system. Future use of head-tracking systems should have automatic correction (re-calibration) and the total weight of the system should be reduced before further testing takes place.

## 8. REFERENCES

- Azuma, R., Bailiot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre B. (2001). Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, 21(6), 34-47.
- Azuma, R., (1997). A Survey of Augmented Reality, *Presence: Teleoperators and Virtual Environments*, 6(4), 355-385.
- Baddeley, A. D. (1976). *The psychology of Memory*. London: Harper & Row.
- Bertelson, P. and Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, 5, 482-489.
- Branch E. J., Sanders, J. D., Kearfott, K. J., Stojadinovic, G., Wehe, D.K. (1999) An Augmented Reality Radiation Display System (ARRDS) for Radiation Protection Applications. *Transactions of the American Nuclear Society*, vol. 81, pp. 247.
- Brooks, F. P. (1996). The Computer Scientist as Toolsmith II, *CACM*, 39(3), 61-68.
- Brooks, J. (1998). SUS - A "quick and dirty" usability scale. (World Wide Web document). <http://www.cee.hw.ac.uk/~ph/sus-html>. Updated 23 October, 1998. Digital Equipment Corporation.
- Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 22, 349-368.
- Buck, J. R., Zellers, S. M. and Opar, M.E. (2000). Control error statistics. *Ergonomics*, 43, 1-16.
- Byers, J., Bittner, A. C. & Hill, S. G. (1989) Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? In A. Mital (Ed.), *Advances in Industrial Ergonomics and Safety I* (pp. 481-485). Taylor & Francis.
- Drøivoldsmo, A., Nystad, E. & Helgar, S. (2001). *Virtual reality verification of workplace design guidelines for the process plant control room (HWR-633)*. Halden, Norway: OECD Halden Reactor Project.
- Drøivoldsmo, A. and Johnsen, T. (2002). Development and testing of an internet-based data collection technique for simulator and real world experimentation. *Manuscript in preparation*.
- Drøivoldsmo, A., Sebok, A. and Nystad, E. (2002). Measuring radiation awareness. *Manuscript in preparation*.
- Dutoit, A. H., Creighton, O., Klinker, G., Kobylinski, R., Vilsmeier, C. & Bruegge, V. (2001). Architectural Issues in Mobile Augmented reality Systems: A Prototype Case Study, *Proceedings of 8<sup>th</sup> Asia-Pacific Software Engineering Conference (APSEC2001)*
- Elzer, P. F., and Behnke, R. (2001). Computer Augmented Reality Techniques for Training, Maintenance, and Repair in Power Plants. *Presented at the OECD EHPG Meeting 2001, Lillehammer, Norway*. Halden, Norway: OECD Halden Reactor Project.
- Feiner, S., MacIntyre B. & Sligmann, D. (1993). Knowledge-based Augmented Reality. *Communication of the ACM* 36(7), 52-62.
- Foxlin, E., Harrington, M., Pfeifer, G. (1998). Constellation™: A Wide-Range Wireless

- Motion Tracking System for Augmented Reality and Virtual Set Applications. *Proceedings of SIGGRAPH'98*, 371-378.
- Gaulin, S. J. & McBurney, D. H. (2001). *Psychology An Evolutionary Approach*. Upper Sadle River, NJ: Prentice Hall.
- Gawron, V.J. (2000). *Human Performance Measures Handbook*. Mahwah, NJ: Lawrence Erlbaum.
- Glass, A.L. & Holyoak, K.J. (1986). *Cognition*. New York: Random House.
- Gravelle, M. D. & Schopper, A. W. (1993). *Measures and methods for characterizing operator performance in nuclear power plant control rooms (CSERIAC-RA-93-007)*. Wright-Patterson Air Force Base, OH: CSERIAC.
- Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.
- Hoff, B., & Azuma, R. (2000). Autocalibration of an Electronic Compass in an Outdoor Augmented Reality System. *Proceedings of International Symposium of Augmented Reality 2000 (ISAR 00)*, 159-164.
- Iguchi, Y. & Rindahl, G. (2002). Development of decommissioning support system using VR Technology. *Proceedings of Japan-Halden MMS workshop* (pp. 74-77). Kyoto, Japan: Symbio Community Forum, Graduate School of Energy Science, Kyoto University.
- Jolicoeur, P. (1992) Identification of disoriented objects: A dual systems theory. In G.W. Humphreys (Ed.), *Understanding Vision* (pp. 181-198). Cambridge, MA: Blackwell.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioural sciences (3rd ed.)*. Pacific Grove, CA: Brooks/Cole.
- Kosslyn, S. E. (1980) *Image and Mind*. Harvard University Press: Cambridge, MA.
- Kosslyn, S. E. (1994). *Image and Brain: The resolution of the imagery debate*. MIT Press: Cambridge, MA.
- Kukulakos, K. N., & Vallino, J., R. (1998). Calibration-Free Augmented Reality. *IEEE Transactions of Visualization and Computer Graphics*, 4(1), 1-20.
- Mann, S. (2000). Telepointer: Hands-Free Completely Self Contained Wearable Visual Augmented Reality without Head-wear and without Any Infrastructural Reliance. *Proceedings of 4<sup>th</sup> International Symposium on Wearable Computers (ISWC2000)*, 177-178.
- Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed reality Visual Displays. *IEICE Transaction of Information Systems*, E77-D(12), 1321-1329.
- Mizell, D. (2001). Boeing's Wire Bundle Assembly Project. In W. Barfield & T.A. Furness (Eds.), *Fundamentals of Wearable Computers and Augmented Reality* (pp. 447-467). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Neumann, U. & You, S. (1998). Natural Feature Tracking for Augmented-Reality, *IEEE Transactions on Multimedia*, 1(1), 53-64.

- Nystad, E., Drøivoldsmo, A. & Sebok, A. (2002). *Use of radiation maps in a virtual training environment for NPP field operators (HWR 681)*. Halden, Norway: OECD Halden Reactor Project.
- Pick, H. L., Jr., Warren, D. H., & Hay, J. C. (1969). Sensory conflicts in judgments of spatial direction. *Perception, & Psychophysics*, 6, 203-205.
- Pretlove, J. R. and Lawson, S. W. (1998). Integrating Augmented Reality and Telepresence for Telerobotics in Hostile Environments, in 'Augmented Reality - placing artificial objects in real scenes', Behringer, Klinker and Mizell (eds), *Proceedings of First IEEE International Workshop on Augmented Reality (IWAR 98), San Francisco, November 1998*, pp. 101-107, published by AK Peters, 1999. ISBN 1-56881-098-9.
- Rauterberg, M., Mauch, T., & Stebler, R. (1996). The Digital Playing Desk: a Case Study for Augmented Reality. *Proceedings of the 5th IEEE International Workshop on Robot and Human Communication*, 410-415.
- Reigstad, M. and Drøivoldsmo, A. (2002). Development of a prototype for augmented reality visualisation of radiation. *Manuscript in preparation*,
- Rekimoto, J. (1997). NaviCam: A Magnifying Glass Approach to Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4), 399-412.
- Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, 143, 594-596.
- Rolland, J., P., Davis, L. D. & Baillot, Y. (2001). A Survey of Tracking Technology for Virtual Environments. In W. Barfield & T.A. Furness (Eds.), *Fundamentals of Wearable Computers and Augmented Reality* (pp. 67-112). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Schwald, B., Figue, J., Chauvineau, E., Vu-Hong, F., Robert, A., Arbolino, M., Schnaider, M., De Laval, B., Dumas De Raully, F., Anez, F. G., Baldo, O., and Santos, J. M. (2001). STARMATE: Using Augmented Reality technology for computer guided maintenance of complex mechanical elements. In *Presented at the e2001 Conference 17-19 October 2001, Venice, Italy*. <<http://vr.c-s.fr/starmate/documents/Publications/e2001Conference17-19October2001VeniceItaly.pdf>>
- Selcon, S. J., Taylor, R. M. & Koritsas, E. (1991) Workload or situation awareness?: TLX vs. SART for aerospace systems design evaluation. *Proceedings of the human factors and ergonomics society 35th annual meeting* (pp. 62-66). Santa Monica, CA: Human Factors and Ergonomics Society.
- Seo, Y., & Hong, K., (2000). Weakly Calibrated Video-Based Augmented reality: Embedding and Rendering through Virtual Camera, *Proceedings of International Symposium of Augmented Reality 2000 (ISAR 00)*, 129-136.
- Shiffrin, R. M. and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Shoval, S., Borenstein, J. and Koren, Y. (1998). Auditory guidance with the Navbelt - A computerized travel aid for the blind. *IEEE Transactions on Systems, Man and Cybernetics*, 28, 259-267.



- Taylor, R. M. & Selcon, S.J. (1991) Subjective measurement of situational awareness. In R.M. Taylor (Ed.), *Situational Awareness in Dynamic Systems (Rep. No. IAM-708)* (pp. 789-791). Farnborough, UK: Royal Air Force Institute of Aviation Medicine.
- Walker, J.T. and Scott, K.J. (1981). Auditory - Visual Conflicts in the Perceived Duration of Lights, Tones, and Gaps. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1327-1339.
- Waller, D. (1999). Factors affecting the perception of interobject distances in virtual environments. *Presence*, 8, 657-670.
- Warren, D. H. (1979). Spatial localization under conflict conditions: Is there a single explanation?. *Perception*, 8, 323-337.
- Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K. & Colucci, D., (2001), High-Performance Wide-Area Optical Tracking – The HiBall Tracking System. *Presence: Teleoperators and Virtual Environments*, 10(1), 1-21.
- Wickens, C.D. & Hollands, J.G. (2000). *Engineering Psychology and Human Performance (3rd ed.)*. Upper Saddle River, NJ: Prentice Hall.
- Yokokohji, Y., Sugawara, Y. & Yoshikawa, T. (2000). Accurate Image Overlay on Video See-through HMDs Using Vision and Accelerometers, *Proceedings of IEEE Virtual Reality 2000*, 247-254.
- You, S., Neumann, U. & Azuma, R., (1999). Hybrid Inertial and Vision Tracking for Augmented Reality Registration. *Proceedings of IEEE Virtual Reality 1999*, 260-267.

## **9. APPENDICES**

Appendix 1: Measurement of radiation awareness

Appendix 2: Subjective rating of situation awareness and workload

Appendix 3: Radiation maps

Appendix 4: Usability questionnaire

Appendix 5: List of acronyms and abbreviations

## 9.1 Appendix 1: Example of questionnaire used for radiation awareness measurement

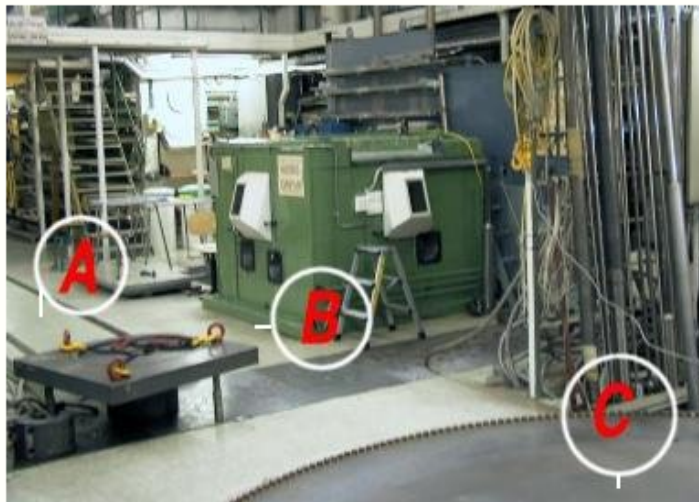
The question presented below is translated from Norwegian.

Where is the highest radiation?

A

B

C



## 9.2 Appendix 2: SART 3-D questions and NASA-TLX

In the test, each question was presented on a separate page (different from the overview presented below). The test questions and descriptions were translated to Norwegian. Each page contained additional instructions for answering the questions.

Demand on attentional resources caused by this scenario	LOW	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	HIGH
Supply of attentional resources demanded in this scenario	LOW	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	HIGH
Understanding of this scenario	LOW	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	HIGH
Situation awareness in this scenario	LOW	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	HIGH

### Overview of questions used as basis for SA measurement

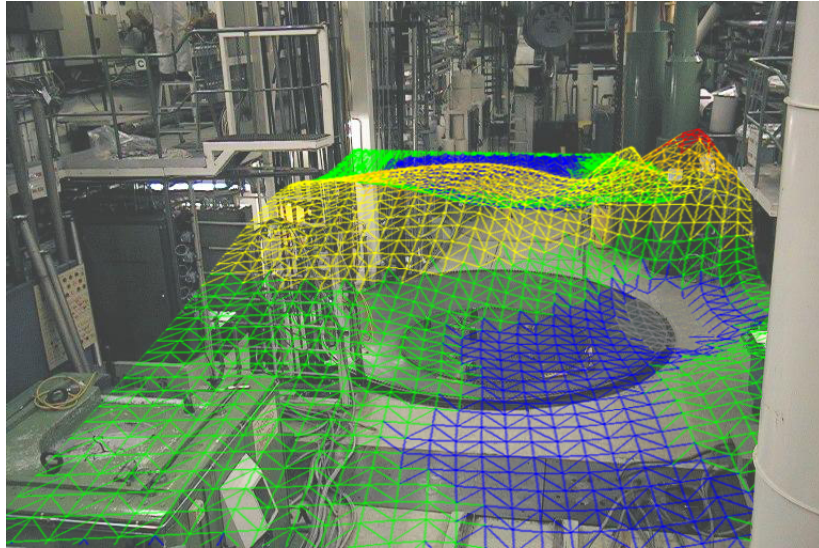
<input type="button" value="Next"/>											
Mental Demands	Low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	High
Physical Demands	Low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	High
Temporal Demands	Low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	High
Performance	Failure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Perfect
Effort	Low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	High
Frustration level	Low	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	High
<input type="button" value="Next"/>											

Page 1 of 5

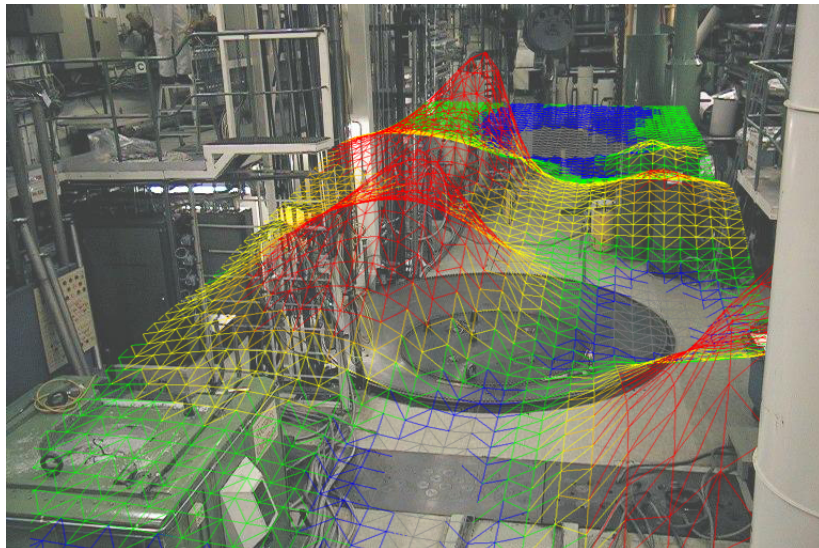
### Overview of the NASA-TLX measure

### 9.3 Appendix 3: Radiation maps

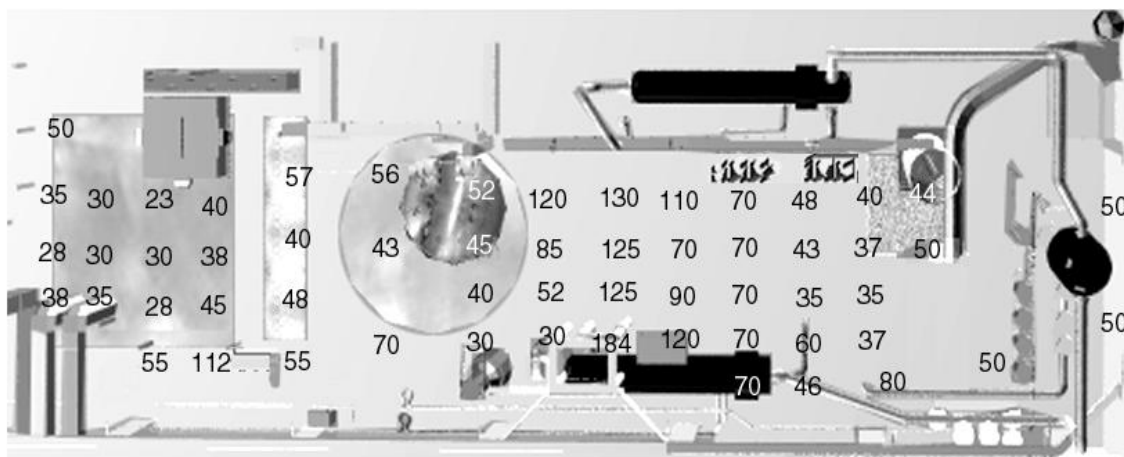
The figures below show the radiation distribution from both scenarios.



*Radiation in scenario 1 as shown in the augmented condition.*



*Radiation in scenario 2.*



*Radiation in scenario 1, as shown in the map condition.*

## 9.4 Appendix 4: Usability ratings

Rating of system usability

1. I needed to learn a lot of things before I could get going with this system

Strongly disagree      1          2          3          4          5          Strongly agree

2. I found the various functions in this system were well integrated

Strongly disagree      1          2          3          4          5          Strongly agree

3. I found the system unnecessarily complex

Strongly disagree      1          2          3          4          5          Strongly agree

4. I felt sure in how to use the system

Strongly disagree      1          2          3          4          5          Strongly agree

5. I think this system will be useful in my daily work

Strongly disagree      1          2          3          4          5          Strongly agree

What did you like about the system?

How would you improve the system?

## 9.5 Appendix 5: Acronyms and abbreviations

AR	Augmented Reality
DOF	Degrees of Freedom
FOV	Field of View
GPS	Global Positioning System
HMD	Head-Mounted Display
HWR	Halden Work Report
ICRP	International Commission on Radiation Protection
IETM	Interactive Electronic Technical Manuals
LCD	Liquid Crystal Display
LED	Light Emitting Diode
$\mu$ Sv	microSievert
mSv	milliSievert
MTO	Man, Technology and Organisation
NASA-TLX	National Aeronautic and Space Administration - Task Load Index
RA	Radiation Awareness
RMS	Root Mean Square
ROE	Reactor Operation and Engineering
SA	Situation Awareness
SART	Situational Awareness Rating Technique
SD	Standard Deviation
STARS	Sticky Technology for Augmented Reality System
ARSVF	Augmented Reality Sound Vibrations First
ARSVL	Augmented Reality Sound Vibrations Last
VNC	Virtual Network Computing
VR	Virtual Reality



# Paper D

Drøivoldsmo, A., Nystad, E. & Helgar, S. (2001). *Virtual reality verification of workplace design guidelines for the process plant control room (HWR-633)*. Halden, Norway: OECD Halden Reactor Project.



# **Virtual reality verification of workplace design guidelines for the process plant control room (HWR-633)**

Asgeir Drøivoldsmo, Espen Nystad and Stein Helgar

## **1. INTRODUCTION**

The aim of the reported work was to validate the Halden Reactor Project (HRP) utilisation of VR technology in the process of verifying control room guidelines in human factors engineering design. In addition to answering the question about validity of the VR model, this report attempts to give an overview of some of the empirical work done in this area. For an easier read and more practical presentation of the results from this experiment, we refer to the short version of the report (Drøivoldsmo et al., 2000).

The importance of a well-designed human-system interface to reliable human performance and process plant safety is widely acknowledged. Review of the control room design plays a significant role in supporting plant safety, and designers rely on the use of human factors engineering guidelines to support the identification of potential safety issues. Criticism is often aimed at the human factors techniques for being unable to give fast and reliable feedback in early stages of the design process. Efficient verification and validation techniques of control room design are therefore in high demand.

Recognised review models for the design process of control rooms (IEC 964, 1989; ISO/FDIS 11064-1, 2000; NUREG-0711, 1994) have a similar general approach to the handling of output from the human factors design verification. They all recommend an iterative process where deviating verification results should be used as input to the appropriate preceding stage or phase in the design process. The ISO human factors programme even proposes a separate intermediate verification to solve all allocation and assignment conflicts before starting the conceptual framework design.

Today, verification of the detailed design is generally done either through a pen and paper verification or verification of a physical mock-up of the proposed control room. In the first case, the verification will usually use the results from analysis performed during the design process, which may be imperfect. For example, the result of a task analysis may be used as a criterion in verifying that all required controls and displays are provided to support human functions (NUREG-0711, 1994; Stubler, O'Hara, Higgins and Kramer, 2000). In the second case, the use of a physical mock-up is expensive in terms of physical infrastructure, in addition to being time- and personnel demanding. For a fuller consideration of HRP human factors design verification and validation methodology, see Collier and Green (1999).

VR technology has proved to be a promising, powerful, and cost effective tool in control room design-work. It enables designers to spend more time evaluating creative new ideas, helping them to identify and eliminate potential problems early in the design process, and lowers the threshold for end-users to more actively participate in the design

process. However, more knowledge should be obtained about the quality of results from utilisation of VR technology in guideline verification.

## **2. HUMAN PERFORMANCE IN VIRTUAL REALITY**

Considering how human beings judge their surroundings, move, and solve problems in virtual environments raises a number of questions. The number one question raised in literature describing virtual environments is the question of presence. Witmer and Singer (1998) described presence as “the subjective experience of being in one place, even when one is physically situated in another”. The perceived meaningfulness of the virtual environment is believed to influence the sense of presence. Hoffman, Prothero, Wells & Groen (1998) found that chess players felt a higher sense of presence when viewing meaningful chess positions than when viewing meaningless positions. Witmer and Singer (1998) argue that presence requires the ability to focus on meaningfully coherent set of stimuli. On basis of the literature, it is therefore believed that the verification of control room design guidelines cannot give dependable results if the virtual environment is not of high enough quality to provide sense of presence.

### **2.1 MOCKUPS**

In NUREG-0711 (1994) a mockup is defined as “a static representation of an human system interface”, as distinct from a prototype which is a dynamic representation of the interface. According to the American Institute of Aeronautics and Astronautics, American standard (1992), mock-ups for hardware development are divided into two-dimensional; three-dimensional static; full scale; and functional full scale (interactive). Using virtual reality adds two more mock-up types to the list; virtual three-dimensional mock-ups may be static or functional. The mock-ups that are compared in the present experiment are a static virtual three-dimensional mockup and a real control room representing a functional full-scale mockup.

### **2.2 VR DEFINITION**

A definition of virtual reality will have to include the system’s ability to represent a physical environment as well as the interactive aspects; the user should be able to navigate in the virtual space or manipulate virtual objects. We will adopt the definition used by Wann and Mon-Williams (1996), where a virtual environment is described as an environment with the following properties:

A virtual environment (VE) provides the user with access to information that would not otherwise be available at that place or time, capitalizes upon natural aspects of human perception by extending visual information in three spatial dimensions and may supplement this information with other sensory stimuli and temporal changes. ... a further constraint to introduce is that a virtual environment enables the user to interact with the displayed data (p. 833).

Successful introduction of VR technology for the purpose of control room design and verification demands technologically simple and inexpensive solutions manageable

without immense investments. Reducing cost by dropping multimodal interaction elements (e.g. tactile sensors, head trackers, etc.) is one way of dealing with this problem. The approach used in the current study is relying on good visualisation and motion for depth cues, and should be inside of the proposed definition by a safe margin. This way of representing the environment is often referred to as desktop VR (Wickens and Hollands, 2000) or Fish-Tank VR (Heim, 1998).

## **2.3 SENSE OF PRESENCE (SOP)**

### **2.3.1 Definition**

Witmer & Singer (1998) defined SOP as "the subjective experience of being in one place, even when one is physically situated in another", or as "experiencing the computer-generated environment rather than the actual physical locale" (p. 225). Regenbrecht, Schubert & Friedman (1998) emphasise the cognitive construction of a space in which the user's body and the virtual environment seems to share the same world. However, as Stanney & Salvendy (1998) and Wilson (1997) points out, there is an absence of agreed methodology and measures for the sense of presence in a virtual environment.

### **2.3.2 Factors affecting SOP**

The literature on SOP mention the following factors as most important for producing a sense of presence: the ability to interact with the VE, user-initiated control, the extent and quality of sensory information, reduction of distracting elements in the physical environment, realism and meaningfulness of the VE, and field of view (Hoffman et al., 1998; Stanney and Salvendy, 1998; Wilson, 1997; Witmer and Singer, 1998). It is of importance that the user is immersed into the simulation so that the user has a sense of being present. The degree of immersion depends primarily on the hardware technology used whereas presence is achieved by making an environment that causes the user to suspend disbelief and accept the computer-generated experience as 'real' (Louka, 1999).

Deisinger, Cruz-Neira, Riedel & Symanzik found screen-based projections to produce a greater sense of immersion in inexperienced users than monitors or head mounted displays (as cited in Stanney and Salvendy, 1998).

### **2.3.3 Presence and performance**

There is no conclusive evidence on the relationship between SOP and performance, though, a number of studies have found correlations between SOP and tracking or search performance (Ellis et al., 1997; Pausch, Proffitt & Williams, 1997, both cited in Stanney & Salvendy, 1998). It has also been suggested that judging the "feel" of simulated places is easier with a greater SOP (Henry and Furness, 1993). However, as Stanney and Salvendy (1998) suggest, it is "...possible that performance and presence do not have a causal relation, but rather are both caused by a third factor, such as improved VEs" (p. 164).

A sense of presence in the VE may induce some kind of "flow" and block out distracting elements in the physical world. Barfield, Zeltzer, Sheridan and Slater (1995) point out the importance of attention; SOP occurs when attentional resources are

allocated to sensory information in the VE. In this case, it is likely that VEs that produce SOP can result in better performance simply because more of the operators' attention is directed to the task at hand.

On the other hand, simulator sickness is negatively correlated with presence, probably because the symptoms of simulator sickness draw attention away from the VE (Witmer and Singer, 1998).

## **2.4 PERCEPTION AND NAVIGATION**

There is probably a close relationship between perception and navigation in virtual environments. If perception (i.e. the ability to construct an accurate representation of the environment) in the VE is poor, then navigation is likely to be poor as a result. The ease of navigation in a virtual environment depends on several factors. Stanney, Mourant and Kennedy (1998) suggest that a measure for navigational complexity "... would likely be dependent on sensorial cues and spatial mediators (e.g. maps) provided by the environment, as well as on the spatial ability of the navigator" (p. 329). Reduced or lacking spatial cues such as shadows, perspective, texture gradients and relative size (Wann and Mon-Williams, 1996) may impair navigation because it makes it difficult to judge one's position relative to the surroundings. Including such cues in the VE facilitates the "natural extraction of environmental dimensions" (Wann and Mon-Williams, 1996, p. 835).

Most VEs use an egocentric point of view where the environment is presented as though the observer were present in and moving through the VE. This can create a keyhole view that makes it difficult to orient because it presents a limited view of the world (Baker and Wickens, 1995). A narrow field of view in real world conditions excludes peripheral information that might be helpful for estimating sizes and distances, Hagen, Jones & Reed (cited in Henry and Furness, 1993), and reduces spatial learning, Sholl (1993), as cited in Waller & Hunt (1998). An exocentric point of view gives an outside-in perspective that includes more of the VE. Using this view instead of an egocentric one has both advantages and disadvantages. The time it takes to search for objects in the VE gets shorter and judgments of relative locations of objects is better with an exocentric point of view compared to an egocentric. But the time it takes to travel toward an object is longer (McCormick, Wickens, Banks and Yeh, 1998).

## **2.5 Estimation of distances and directions in VEs:**

People generally underestimate distances both in the real world and in virtual environments, but underestimations are usually greater in VE than in the real world (Henry and Furness, 1993; Witmer and Kline, 1998; Witmer and Sadowski, 1998). Richardson, Montello and Hegarty (1999), found no differences for distance estimations between real world and VE. However, according to Richardson et al., subjects navigating in a VE acquire a poorer spatial representation of the environment than subjects navigating in the real world.

A factor influencing the spatial perception is Geometric field of view (GFOV), representing the visual angle depicted in the virtual scene. Waller (1999) found that a GFOV of 80 degrees gave more accurate estimates of exocentric distances than a GFOV

of 50 or 100 degrees (exocentric distance: distance between two objects in the environment – in contrast to egocentric distance that is the distance between the observer and an object in the environment). Waller suggests that "the effect of GFOV on accuracy is due to the magnification (or minification) induced by lower (or higher) GFOV" (p. 666), or the fact that a narrow GFOV makes it difficult to see both objects in the picture at once. GFOV also affected the time to complete the estimates, with narrow GFOV taking longer than wide GFOV

The reason for the greater underestimations of distances in VEs is thought to be the lack of depth cues like motion parallax, the absence of additional features in the environment (Witmer and Kline, 1998), narrow vertical and horizontal GFOVs (Henry and Furness, 1993; Witmer and Kline, 1998; Witmer and Sadowski, 1998), height in the visual field, convergent linear perspective, and relative size (Witmer & Sadowski, 1998).

## **2.6 SIMULATOR SICKNESS**

Simulator sickness is probably caused by a mismatch between the sensory stimuli felt by the body and the stimuli the organism expects based on past experience (Stanney and Hash, 1998). This mismatch cause illness that manifest itself in nausea and vomiting, eye strain and disorientation, Kennedy, Berbaum and Drexler (as cited in Stanney and Hash, 1998). Held and Reason concluded in their studies that it is possible to adapt to motion sickness if users have control over their movements and receive a sensory response to their actions (as cited in Stanney and Hash, 1998). Simulator sickness seems to decrease with repeated exposure to the VE, probably because of adaptation. But it increases with prolonged exposure to the environment (Kennedy, Stanney & Dunlap, in press, cited in Stanney and Salvendy, 1998). Thus, short repeated exposures seem to be best for avoiding simulator sickness.

Stanney and Hash (1998) found that an active-passive control of the situation leads to lower levels of motion sickness than either passive or active control conditions. In the active-passive condition, subjects had control over movements that were appropriate for the task, while other movements were restricted or eliminated. The subjects were thus prevented from receiving excessive sensory information that could lead to effort to stabilize their movements, resulting in motion that could conflict with the visual scene. Control of, and feedback from, the VE is also an important factor. Lag between the motions the user makes and when the display is updated may have an effect on simulator sickness (Stanney et al., 1998).

Other negative effects that may occur during or after exposure to a VE are disturbed locomotor and postural control, perceptual-motor disturbances, drowsiness and fatigue, and lowered arousal or mood (Stanney and Salvendy, 1998).

## **2.7 DISPLAY FACTORS**

The VE is presented to the viewer through a physical display, and it is important to be aware of the effects of different display characteristics. Two key characteristics are the frame rate and the display type, which may be headmounted (HMD), desktop, or projected.

The minimum frame rate recommended for VE presentation seems to have increased as a function of better hardware over the last decade. In their study from 1989, Zeltzer, Sturman & Pieper (as cited in Baker and Wickens, 1995) found that with frame rate of 6-10 frames per second (Hz), the VE felt natural and effective. However, at 3 Hz it was difficult or impossible to use. In the early 1990s, minimum acceptable frame rates of 6 and 7 Hz were found by Airey, Rohlf, and Frederick (1990); and Pausch (1991), respectively (both as cited in Watson, Walker, Ribarsky and Spaulding, 1998). Card, Robertson and Macinlay (1991); McKenna and Zeltzer (1992), and Bryson (1993) all recommended 10 Hz as minimum frame rate (all as cited in Watson et al., 1998), while interactivity appeared to improve significantly with frame rates above 17 Hz (Watson et al., 1998). In line with this finding, Reddy (1997) found a sharp improvement of performance as frame rate increased up to 10-15 Hz, and a substantially less rapid increase after 15 Hz. Variation in system response is another important factor. Watson et al. (1998) found standard deviation system response (SDSR) time to have an influence on performance, but only for SDR of 82 ms or more.

On most performance measures, Ye, Banerjee, Banerjee and Dech (1999) found no significant differences between a desktop VE and a projected Cave Automatic Virtual Environment (CAVE™), where the environment is projected on to four walls surrounding the viewer, and gives a 3D-effect using alternate left- and right-eye projections that are seen through a pair of LCD shutter glasses. Henry and Furness (1993) found a difference in distance estimations for a head-tracked HMD (which lead to greater underestimations) compared to non-tracked HMD or a desktop display, but no significant difference between the desktop display and the non-tracked HMD. As previously mentioned, a projected display was found to produce the greatest sense of presence (Deisinger et. al 1997, as cited in Stanney and Salvendy, 1998).

The overall objective of this research was to determine how accurate verification of guidelines for different tasks could be done in a VE compared to a real environment. If guideline verification in a VE should be valid, then the performance in either of the two conditions should be the same.

## **3. METHOD**

### **3.1 Experimental design**

Each of twelve subjects performed a guideline verification of a control room in the Halden Boiling Water Reactor control room (HBWR condition) and in a VR model of the same control room (VR condition), verifying a set of randomly chosen guidelines from five different sections of chapter seven in NUREG-0700 (1996). Thus, a within-subjects design was used where the independent variable was the type of control room used for guideline verification, and the dependent variables were the rating on each guideline and the time used on the verification. The design was counter balanced by letting half of the subjects perform the HBWR control room guideline verification first, while the other half performed the VR model verification first. To prevent learning effects, the second guideline verification took place approximately two months after the first. This time interval was assumed long enough to prevent subjects on the second run



from remembering how they rated guidelines the first time. To achieve a high degree of realism in the guideline verification, all five tasks were presented in the same order as they appear in the NUREG, i.e., all 50 guidelines were presented in the same order for all subjects in both conditions. This order of presentation is the natural way of utilising the NUREG software tool. Balancing out carry over (practice) effects from the tasks was believed to introduce an artificial variable in the study.

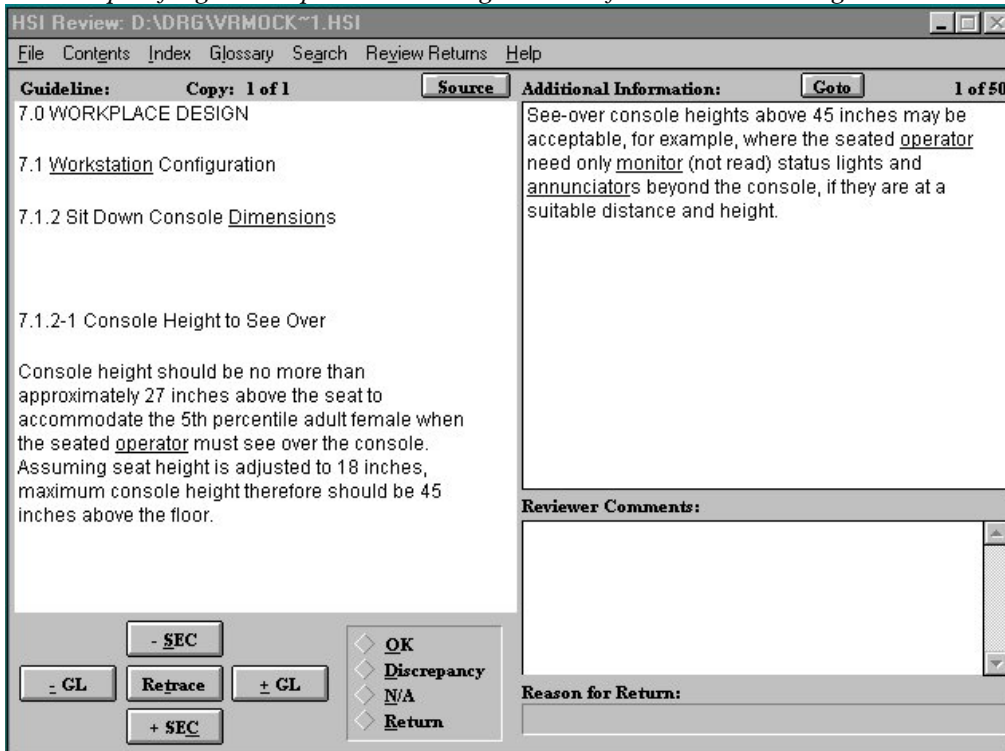
### **3.2 Subjects**

The participants were drawn from the population of potential practitioners of control room verification and validation systems at the Halden Project, i.e. the staff from the Man Machine Systems Research division and the Control Room Systems division. To achieve a statistical power of approximately .8, the minimum number of subjects was estimated to be approximately 12. A total of 15 subjects were needed to get 12 complete data sets. Two subjects completed only half of the guidelines, and one subject withdrew from the experiment because of simulation sickness caused by the VR model. The 12 who completed the experiment had a mean age of 34, ranged from 26 to 55. Two were female. The three subjects who did not complete had a mean age of 46. One of them was a female. All subjects received information about the purpose of the experiment in advance and signed a consent form before beginning the experiment.

### **3.3 Materials**

Five sections from chapter 7 of NUREG-0700 (1996) that were relevant to the HBWR control room were identified, and then 50 guidelines were randomly selected from those sections. These 50 guidelines were presented to the subjects on a Compaq Armada 1130T laptop PC (120MHz Pentium, 32MB RAM) using the software version of NUREG-0700, version 1.0. The laptop and software were used in the same way in both the HBWR and the VR condition to present and record how well the control room conformed to each guideline (see Figure 1).

Figure 1. Example of a guideline presented in the guideline software tool. The navigation buttons in the



lower left were used for paging through the guidelines, and the radio buttons were use to score the guidelines.

In the HBWR condition subjects performed measurements and evaluations in the actual HBWR control room. They used a Bosch DLE 30 laser range finder and a standard tape measure to measure distances.



Figure 2. Experimental set-up in the VR condition: The subject is placed in front of a screen showing the projection of the VR model.

In the VR condition subjects used a VR application where they could navigate in a VR model of the HBWR control room using a mouse and keyboard. Distances were measured by selecting two points (locations) in the model using the mouse. A fixed-dimension view cone (the conical area that can be seen by looking at some direction) and a manikin representing a control room operator were available in the VR model to

view the operator's field of view. The VR application was presented using a SONY multiscan projector VPH-G70 on a screen 2,6 m from the subject, showing a picture of size 3,0 m x 2,3 m with a resolution of 1280x1024 pixels and a refresh rate of 85 MHz (see Figure 2). The VR application ran on a Pentium III PC (667 MHz, 256 MB RAM) with a "Creative 3D Blaster GeForce 256 Annihilator PRO" graphics card (32 MB SGRAM), giving a minimum refresh rate of 20 frames/s. The model of the control room had a geometric field of view of 50 degrees. The subjects received a map of the control room before starting the experiment to get an overview of the layout of the room.

The VR model was built in Multigen Creator version 3.2, consisted of approximately 5000 polygons and 3.1 MB of textures in two different sizes, 512x1024 pixels and 256x512 pixels. The model was presented using a Java3D VRML97 loader in a specially designed software tool.

### **3.4 Procedure**

At the first experiment session, each subject was first asked to sign the consent form and to fill in a questionnaire for demographic data.

#### **HBWR condition**

The experiment in the HBWR condition took place during normal working periods in the reactor control room, with control room staff present. This was due both to the resources of the project and to the wish to create an experimental situation that was similar to the situation in which guideline verification usually takes place.

They were then instructed in using the laser range finder and the NUREG-0700 software. Subjects were told to rate the control room using one of four categories: If the control room conformed to the guideline, it should be rated 'OK'. If it did not conform, it should be rated 'Discrepancy', and subjects were asked to write a short description of the discrepancy in the software. If the guideline was not applicable to the control room, it should be rated 'N/A', and if it could be rated only if the subject received additional information or other measurement instruments, the subject should use the category 'Return' and choose which of the following reasons were most appropriate: 'Need operator input', 'Need plant management input', 'Requires measurement instrument', 'Requires additional documentation' or 'Other requirement'. The subjects were then told that there were 50 guidelines to be tested, and that they should work efficiently, but only so fast that they were certain to verify the guideline correctly. If they had problems using the guideline software, they could get help from the experimenter. For guidelines where the subjects were asked to look at labelling, they were instructed to look only at the leftmost stand-up panel.

#### **VR condition**

The subjects first received a map of the control room, and they were then presented a short tour of the VR model by the experimenter. They were then given instructions for how to navigate, measure distances (as shown in Figure 3), and view the operator's field of view in the model. They were also allowed to practice navigating and using the measurement tools and were asked to do two navigation tasks to show that they could

use the navigation tools. The subjects in the VR condition were given the same instructions as in the HBWR condition in using the NUREG-0700 software, and regarding how fast they should work and getting help from the experimenter. After completing the verification, they were asked to fill in a questionnaire for evaluating the VR application.



Figure 3. Screen dump of measurement in the VR condition. The length of line drawn from the manikin's head to the control panel (2,62 meters) is displayed in the upper right part of the figure.

### 3.5 Data collection

The subjects were timed from the start to the end of the guideline verification session. The demographic data collected were age, sex, education and work experience and prior experience with human factors guidelines work, NUREG-0700 and VR applications (for the full questionnaire, see appendix A). After the VR session, subjects evaluated the usability of the VR application, rating the ease of use and complexity, what they liked and disliked and what they would improve (for the full questionnaire, see appendix B). The data log from the NUREG-0700 software contained the ratings for each of the 50 guidelines including the subjects' comments.

## 4. RESULTS AND DISCUSSION

### 4.1 Main analysis

Subject's ratings were compared to an instructor's copy based on detailed measurements from the HBWR control room, consensus between several independent scorers in the experimental staff and consultation of independent experts for ambiguous guidelines. The reason for 'Return' was not taken into consideration in the overall analysis. The analysis is based on scores in percentage of correct answers for each of the five tasks in both control room representations.

The five sub-categories based on the classification used for the NUREG-0700 chapter sections are described in Table 1 and appendix C.

*Table 1 Categories of guidelines with respective correct scores in percent (N=12).*

Categories of guidelines	HBWR M and SD	HVRC M and SD
Workstation console and desk, covering sit down console and desk dimensions (15 guidelines)	56.6 (.09)	51.1 (.12)
Workstation chair (5 guidelines)	66.6 (.13)	46.6 (.18)
Control room configuration including ambience and comfort (13 guidelines)	73.7 (.17)	60.9 (.14)
Panel layout (5 guidelines)	33.3 (.27)	35.0 (.21)
Panel labelling (12 guidelines)	81.9 (.10)	61.1 (.16)

A 2 x 5 within subjects analysis of variance on performance was conducted to examine the effect of the factor control room representation (two levels – HBWR and HVRC) in all five tasks. The analysis showed a significant main effect of control room representation,  $F(1,11) = 37.0, p < .01$ .

No interactions were found between task and control room representation and order of presentation, but there was a main effect of task  $F(4,80) = 18.4, p < .01$ . Further analyses with Sheffé post hoc tests, showed that the panel layout task was different from all the other four tasks ( $p < .01$ ), and that the console task differed from the panel labelling task ( $p < .05$ ).

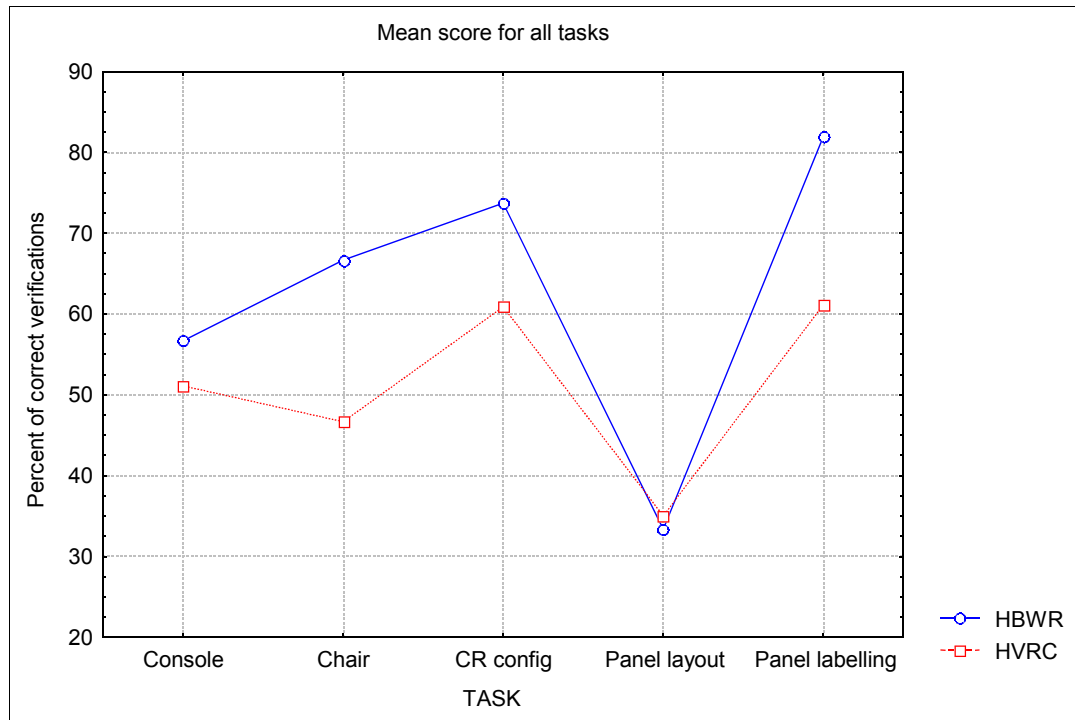


Figure 4. Mean score of correct verification as a function of task and experimental condition

As shown in *Figure 4*, results were dependent on the task performed. In the workstation chair task (task 2) and the control room configuration task (task 3) the HBWR condition showed marginally better performance ( $F(1,11)=6.6$ ,  $p<.05$  and  $F(1,11)=7.53$ ,  $p<.05$  respectively). The panel-labelling task showed a clear superiority for the HBWR condition ( $F(1,11)=14.74$ ,  $p<.01$ ). For the remaining two tasks, no difference was found between the two conditions.

An analysis of variance was performed to determine possible effects on performance from order of presentation. No significant effects were found, suggesting that the time delay was successful in preventing learning effects.

## 4.2 Workstation console and desk

Task number one included a number of detailed measurements. The VR environment provided special measurement tools, designed for measurement on the accuracy level needed for solving this task, and these tools seem to have been sufficient with respect to usability and accuracy. The data did not allow a detailed comparison of the different tools used within the task, but the usability evaluation (see section VR verification system usability) pointed out some room for improvement.

What is important regarding the measurement tools is that the subjects were able to select the correct points to measure. This means they were able to see where to click the measurement tool. Waller (1999) suggests that inaccurate distance judgments with low geometric fields of view (GFOV) happen because it is less likely that two objects are displayed together, so the distance estimation task is inferential rather than perceptual. The GFOV in this experiment was due to technical limitations set to 50 degrees. The distance measurement tool prevented a negative effect, but a small key hole view makes

it difficult to see if there are other objects/elements in the room that are relevant for the rating. Changing the GFOV depending on the need for detailed or more general overview would probably be a better solution and should be considered for the future.

### **4.3 Chair and panel labelling**

For tasks two and five, the guideline verification of the control room chairs and panel layout, better performance was observed in the HBWR condition than in the HVRC condition. The reason for this difference between the control room representations could be connected to the level of detailed modelling and functionality of the VE. The simulation of the chairs did not provide the functionality needed for proper evaluation, i.e., turning and rolling of the chairs. The panel-labelling task included a number of guidelines where font sizes, space between letters etc, needed to be verified. The resolution of the VR environment used in the experiment made it impossible to judge single letters.

### **4.4 Panel layout**

No difference was found between the control room representations for this task. The low performance level reflects a low inter scorer agreement among the subjects in both the HBWR and the HVRC condition. The nature of the guidelines drawn for this task was rather open for judgement and not particularly easy to apply on the control panel selected for verification.

### **4.5 Control room configuration**

For the control room configuration task, the reason for better performance in the HBWR condition is unclear. Several of the participants reported feeling unwell during the session in the VR environment. According to Witmer and Singer (1998) simulator sickness is negatively correlated with sense of presence, probably because the symptoms of simulator sickness draw attention away from the virtual environment. However, the limited records of the sickness problem did not allow a conclusion in this study.

In the study of distance estimations by Witmer and Sadowski (1998), there was a negative learning effect for those who first trained in the VE. These subjects made worse distance estimations in the real world condition than those who first trained in real world. No such effect could be detected in the present study. Performance was equal for those who first made the guideline ratings in VE, and those who first made the ratings in the real control room. As mentioned above, the distance measurement tool had a positive effect for the distance estimation tasks in VR, and it is probable that this prevented the type of effect found by Witmer & Sadowski.

The nature of this task implies a need for an overview of the control room. Screen-based projection has been found to produce a greater sense of presence in inexperienced users than monitors (Deisinger et. al 1997, as cited in Stanney and Salvendy, 1998). The use of a large screen in this experiment should give good performance conditions in the VR environment. According to the literature, system responsiveness should be satisfactory for good performance (Watson et al., 1998). What remain unexplored are the navigation tools (see system usability, in the next section) and effects of fatigue.

Stanney and Salvendy (1998) have described drowsiness and lowered arousal or mood during exposure to virtual environments. The fact that the control-room configuration task on average was after about half an hour of exposure to VR could have had a negative effect.

It should also be mentioned that one of the guidelines comprising this task was difficult to interpret from the VR model. This particular guideline in question required a check of whether procedures could be removed from a bookshelf and placed on the operators desk. The VR model did not allow this check to be performed, but this guideline alone does not by explain the performance difference between the two conditions.

#### **4.6 VR verification system usability**

Response times for the verification session were examined by a 2 (test order, between subjects) x 2 (control room representation, within subjects) ANOVA. There was no difference in time spent for the validation between the two control room representations, but an order effect was found. The main effect of test order,  $F(1,10)=7.69$ ,  $p < .05$ , suggest that the subjects who performed the verification in the HBWR condition first, spent more time in the HVRC condition than subjects that did the verification in the opposite order. However, this did not result in poorer performance in the guideline verification. Looking into possible explanations demanded a more thorough evaluation of the questionnaire data.

The usability evaluation revealed an overall slightly positive rating of the VR software tool. However, on the five item scale questionnaire, the group that performed the verification in the HVRC condition first, rated the software tool higher ( $M=3.4$ ,  $SD=0.4$ ) than the group that performed the HBWR verification first ( $M=2.4$ ,  $SD=0.7$ ),  $t(10)=2.73$ ,  $p<.05$ . Examining the individual questions, the HBWR-first group found the system more complex ( $M=4.7$ ,  $SD 0.5$ ) vs. ( $M=3.7$ ,  $SD=0.8$ ),  $t(10)=2.54$ ,  $p<.05$ , and the system less well integrated ( $M=4.3$ ,  $SD=0.8$ ) vs. ( $M=2.8$ ,  $SD=1.0$ ),  $t(10)=2.9$ ,  $p<.05$ ) than the other group. Looking at the open comments revealed that the complexity problems in the HBWR first group could be connected to problems with the navigation and trouble using the measurement tool. For this group 4 out of 6 subjects complained about problems with the navigation while this topic not were specifically mentioned by any of the HVRC first group members.

In a VE without collision detection, like the one used in the present study, poor navigation could lead to several time-consuming problems. Getting lost outside the control room and attempts to use tools from behind transparent surfaces (e.g., doors with glass) were two frequently observed problems in this study. Unintentionally moving into walls, panels and rooms adjacent to the control room were also a problem for several subjects. It is reasonable to believe that these navigation problems could have resulted in more time spent among the subjects that found navigation difficult than it did for the subjects not reporting navigation problems. If this conclusion is correct, the extra time spent by the HBWR first group in the HVRC condition was a technical problem, and not any transfer effect from being in the HBWR first. As long as the subjects were instructed to do the validation at their own pace, these navigation problems should not necessarily degrade the correctness of the validations. Increasing the number of landmarks was mentioned as a possible improvement of the VE user



interface for navigation, but it would probably be better to try to avoid the problem by enabling collision detection and increase the GFOV.

The most striking results from the open usability questions was the fact that 10 out of 12 subjects named the resolution of the panels in the model as problematic. This gives an explanation to the poor performance in the VR panel labelling task.

## **5. GENERAL DISCUSSION**

### **5.1 The importance of visualisation**

Decreasing cost and less complexity are good technological arguments for more use of 3D visualisation in any design process. However, the criterion for successful use of VR in work with design guidelines verification is whether techniques adapted to review regulations can be useful. The best argument for use of a visualisation tool as a VR-model in the design verification process is participation from a broader field of expertise. One of the best resources for critical reviews of designs is the plant operator. Technical documents like task analyses and function allocation descriptions are not always easily interpreted by non-experts, and visualisations are then the best means of communication of potential design problems. Starting the visualising of initial ideas early in the design and implementation process, based on vendor solutions and operating experience reviews, could be a useful tool for an early verification stage as suggested in ISO 11064 (1989).

### **5.2 Other industries**

In this report the focus is on the rather comprehensive guidance outlined in NUREG-0711 (1994) and IEC 964 (IEC 964, 1989) developed for the nuclear power plant control room design verification. For process industries in general the regulations for HSI design verification is not as extensive as for the nuclear domain, but the user's need for good human system interface design is of course the same, and there is increasing focus on this area. Looking into the offshore petroleum industry for example, the regulations for control room design reviews are becoming more extensive, and a good guess is that they will closely resemble the nuclear regulations in a few years.

### **5.3 Further development and testing**

Based on the results from this experiment, it is recommended to initiate a new phase of validation. Partial rebuilding of the VR-mockup used for this study should be done to make a new test where the textures are modelled and replaced with objects. This will make verification of labelling guidelines possible.

The inferiority of the VR-model in the panel labelling task showed a need for new measurement tools supporting more reliable judgement of text size. Such tools will make possible exact calculation of viewing distance and letter size recommendations in hierarchical labels. Further development of the tool visualising angles of view is also recommended.

The next phase of development should take the results from the control room layout task into close examination. Better depth cues could be provided through more accurate lighting in the model. More accurate light settings can provide more information through shadows and better colour representations.

## **6. CONCLUSION**

The results from this study indicate that satisfactory verification of guidelines can be found only if the VR environment provides good support for the evaluator. An unexpected difference found in performance between the two experimental conditions in the control room configuration task, indicates a need for a greater sense of presence in the VR model. Even if overall performance seems higher in the real control room, much of this difference is due to the resolution of the model in this experiment. One should also bear in mind that the situation facing the evaluator in a design mock-up could never be as good as in the real control room condition used in this experiment.

## 7. REFERENCES:

- American National Standard. (1992). *Guide to human performance measurement* (BSR/AIAA G-035). Washington DC: American institute of Aeronautics and Astronautics.
- Baker, M.P. & Wickens, C.D. (1995). Human factors in virtual environments for the visual analysis of scientific data. Technical report 032. . NCSA.
- Barfield, W., Zeltzer, D., Sheridan, T. & Slater, M. (1995) Presence and performance within virtual environments. In W. Barfield & T.A. Furness (Eds.), *Virtual environments and advanced interface design* (pp. 474-513). New York, NY: Oxford University Press.
- Collier, S.G. & Green, M. (1999). *Verification and validation of human factors issues in control room design and upgrades* (HWR-598). Halden, Norway: OECD Halden Reactor Project.
- Drøivoldsmo, A., Helgar, S., Nystad, E., Johnsen, T., Louka, M.N. & Seregi, L. (2000) Verification of workplace design guidelines for the control room: Can virtual reality replace the physical mock-up? In K.I. Fostervold & T. Endestad (Eds.), *Proceedings of Nordiska Ergonomisällskapetets Årskonferanse, 2000* (pp. 73-77). Stockholm, Sverige: Nordiska Ergonomisällskapet.
- Heim, M. (1998). *Virtual realism*. New York: Oxford University Press.
- Henry, D. & Furness, D. (1993) Spatial perception in virtual environments: Evaluating an architectural application. *Proceedings of IEEE 1993 Virtual Reality Annual International Symposium, VRAIS '93* (pp. 33-40). Piscataway, NJ: IEEE Service Center.
- Hoffman, H.G., Prothero, J., Wells, M.J. and Groen, J. (1998). Virtual chess: Meaning enhances users' sense of presence in virtual environments. *International Journal of Human-Computer Interaction*, 10, 251-263.
- IEC 964. (1989). *Design for control rooms of nuclear power plants* (International Standard 964 (1989-3). Geneva: International Electrotechnical Commission.
- ISO/FDIS 11064-1. (2000). *Ergonomic design of control centres - Part 1: Principles for the design of control centres* (ISO/FDIS 11064-1). Geneva: ISO.
- Louka, M.N. (1999). *An introduction to virtual reality technology* (HWR-588). Halden, Norway: OECD Halden Reactor Project.
- McCormick, E.P., Wickens, C.D., Banks, R. and Yeh, M. (1998). Frame of reference effects on scientific visualization subtasks. *Human Factors*, 49, 443-451.
- NUREG-0700. (1996). *U.S. Nuclear Regulatory Commission. Human-System Interface Design Review Guideline* (NUREG-0700, Rev. 1, Vol. 1). Washington, DC, USA: The NRC Public Document Room.

- NUREG-0711. (1994). *U.S. Nuclear Regulatory Commission. Human Factors Engineering Program Review Model* (NUREG-0711). Washington, DC, USA: The NRC Public Document Room.
- Reddy, M. (1997). *The effects of low frame rate on a measure for user performance in virtual environments* (Technical Report ECS-CSG-36-97). Edinburgh, Scotland: Department of Computer Science, University of Edinburgh.
- Regenbrecht, H.T., Schubert, T.W. and Friedmann, F. (1998). Measuring the sense of presence and its relations to fear of heights in virtual environments. *International Journal of Human-Computer Interaction*, 10, 233-249.
- Richardson, A.E., Montello, D.R. and Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory and Cognition*, 27, 741-750.
- Stanney, K. and Salvendy, G. (1998). Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda. *International Journal of Human-Computer Interaction*, 10, 135-187.
- Stanney, K.M. and Hash, P. (1998). Locus of user-initiated control in virtual environments: Influences on cybersickness. *Presence*, 7, 447-459.
- Stanney, K.M., Mourant, R.R. and Kennedy, R.S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence*, 7, 327-351.
- Stubler, W.F., O'Hara, J.M., Higgins, J.C. & Kramer, J. (2000). *Human systems interface and plant modernization process: Technical basis and human factors guidance* (NUREG/CR-6637). Washington, DC, USA: U.S Nuclear Regulatory Commission, Office of Nuclear Regulatory Research.
- Waller, D. (1999). Factors affecting the perception of interobject distances in virtual environments. *Presence*, 8, 657-670.
- Waller, D. and Hunt, E. (1998). The transfer of spatial knowledge in virtual environment training. *Presence*, 7, 129-142.
- Wann, J. and Mon-Williams, M. (1996). What does virtual reality NEED?: Human factors issues in the design of three-dimensional computer environments. *Int j hum-comp stud*, 44, 829-847.
- Watson, B., Walker, N., Ribarsky, W. and Spaulding, V. (1998). Effects of variations in system responsiveness on user performance in virtual environments. *Human Factors*, 40, 403-414.
- Wickens, C.D. & Hollands, J.G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wilson, J.R. (1997). Virtual environments and ergonomics: needs and opportunities. *Ergonomics*, 40, 1057-1077.
- Witmer, B.G. and Kline, P.B. (1998). Judging perceived and traversed distances in virtual environments. *Presence*, 7, 144-167.

- Witmer, B.G. and Sadowski, W.J. (1998). Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors*, 40, 478-488.
- Witmer, B.G. and Singer, M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7, 225-240.
- Ye, N., Banerjee, P., Banerjee, A. and Dech, F. (1999). A comparative study of assembly planning in traditional and virtual environments. *IEEE Transactions on Systems, Man and Cybernetics*, 29, 546-555.



## Paper E

Drøivoldsmo, A. (2000) Study of errors by means of simulation and training. In P.F. Elzer, R.H. Kluwe & B. Boussoffara (Eds.), *Human error and system design and management* (pp. 201-206). London: Springer-Verlag.





# Study of Errors by Means of Simulation and Training

Asgeir Drøivoldsmo

OECD Halden Reactor Project, P. O. Box 173, N-1751 Halden, Norway,

Email: asgeird@hrp.no

**Abstract:** The ongoing Human Error Analysis Project (HEAP) was initiated within the OECD Halden Reactor Project in 1994. Its objectives are to develop a better understanding and explicit model of how and why cognitive errors occur, and to provide design guidance to avoid, or compensate for, cognitive errors. During the project period, results have led to practical insights concerning method development, and investigation of, simulation as a way of investigating how operators diagnose cognitively challenging scenarios.

## Introduction

A main problem, in the study of complex operating situations, is that it is not clear what should actually be measured. In the special case of a program studying “human error,” one important question is how to find good measurable indicators describing the origin of human erroneous actions. This question is independent of whether one chooses to use classification<sup>1 2</sup>, deeper qualitative analysis, or more general performance indicators like, i.e., operator performance<sup>3</sup>; plant performance<sup>4</sup>; situation awareness<sup>5 6</sup> as the operational tools for investigation of the topic.

Fortunately, without entering the ambiguous variety of definitions and classifications of human error, one can agree that a basis for this research lies in collection of good and reliable data with a potential for aggregation upon a level where meaningful analyses can take place. One way of acquiring such data is to use simulations. Data are then depending on a realistic simulations and tasks, and a representative population of operators (subjects). It is important to include these requirements to comply with the validity criteria that should guide methodologies in this domain<sup>7 8</sup>.

This article will describe the HEAP application of simulation as a tool for gathering good and valid data with the purpose of studying human erroneous actions.

## Halden Man Machine Laboratory

The objectives of HEAP are addressed by an experimental programme, carried out in the Halden Man Machine Laboratory (HaMMLab). The main tool in HaMMLab is the Nokia Research Simulator (NoRS). The NoRS plant model simulates a pressurised water reactor power plant with two parallel feedwater trains, turbines and generators. It is closely linked to the plant model used in the large-scale simulator at the Loviisa nuclear power station in Finland. HaMMLab is different from a standard training

simulator with respect to its infrastructure. In addition to the opportunity to run realistic power plant scenarios, the laboratory has been equipped for flexible experimental set-up, detailed observation and data registration of operator problem solving behaviour.

## **Operators and Scenarios**

The operators used in HEAP experiments have been professional operators from the Loviisa plant and specially trained operators from the Halden Boiling Water Reactor (HBWR). Dependent on the purpose of the study, the choice of operators is critical. Results from Follesø et al.<sup>9</sup> showed that, in cognitively demanding scenarios, these two groups of operators differed significantly in problem solving style and performance. Lessons learned from pilot experiments in HEAP is, that professionally trained operators with long experience, can not be directly compared with crews made up of trained operators from a research reactor.

Design of scenarios in HEAP has been guided by the aim of the project. The objective has been to study cognitive problem solving and the focus has been on diagnostic scenarios with little or no use of operating procedures. The practical aspects of operation must be handled with care trying to test theoretical hypothesis. Here the researcher must rely on subject matter experts (SME) with experience from the simulated process. Safety culture and operational goals should be taken into account. HEAP has supported the investigation of, safety vs. economical goal conflicts, with the purpose of developing a tool to aid the SME in the scenario profiling<sup>10</sup>. Another tool developed is an eight-factor complexity profiling questionnaire, which has proven effective for performance prediction<sup>11 12 13</sup>.

## **Data Collection**

It has been essential to gather enough data about each individual subject to identify what information the individual has, and how he is processing it. A logical starting point for data collection has been verbal data. Verbal data provides a high output rate of information about the operators' attention<sup>14 15</sup>. Used with care, verbal protocols are powerful tools for insight into problem solving<sup>16</sup>.

Four different types of verbal data have been investigated. These four are operator concurrent verbal protocols, operator interrupted verbal protocols, operator retrospective verbal protocols (auto-confrontation) and SME concurrent verbal protocols<sup>17 18 19 20</sup>.

The evolution of new and better technology for EMT (Eye Movement Tracking) has given the opportunity to use this method in HaMMLab experiments without interfering with the operators' problem solving. This is a very powerful source of information about the operators' voluntary visual information gathering, and combined with verbal protocols, EMT can contribute to better accuracy<sup>21 22 23</sup>. Promising results from EMT analysis have also been achieved in the development of measures of tracing cognitive activity<sup>24</sup>, and<sup>25</sup>.

## **Scoring and Analysis of Performance Data**

One important lesson learned in HEAP is the need for a close co-operation with SME throughout the whole scoring and analysis phase of an experiment. There is no way the

complex process like the one found in a nuclear power station can be sufficiently learned and understood of research staff without hands on experience. Scoring and analysis where process knowledge comes into consideration should always be guided by subject matter expertise.

A common goal for the data collection, scoring and analysis of data in HEAP has been to make the performance scores comparable across scenarios, and sensitive to different levels of problem complexity. To achieve flexibility in the scoring, and to meet different demands for details, data have been recorded on the highest practical resolution and attached to a common timeline for all data sources.

The cost of data gathering implies that one will always have to live with a limited number of subjects<sup>26</sup>. Data scoring connected to the time line allows expanding the use of within subject design, through use of time-windows based on important events or stages in the scenarios. Within-subject designs used across scenarios have also proven useful in the analysis in situations where effects are scenario dependent, although interpretation can be difficult when such explorative techniques are employed<sup>27</sup>.

## **Conclusions**

HaMMLab experiments are better described as realistic simulations applying a set of experimental control mechanisms, than traditional psychological laboratory experiments carried out in a nuclear power plant simulator. The experimental methodology developed in HEAP has aimed at data collection and meaningful analysis within the natural context of simulation, preserving the access to the powerful tools of experimental designs.

At the same time as the hypothetical – deductive approach has been the central paradigm in most HEAP experiments, HEAP can be said to be inductive in its underlying research strategy. Technical development allows more and faster data registration and analysis. Tracing both the operator and the process together, through data from e.g., eye-movement tracking, operator activity logs, process logs, audio and video, have opened a new path into the study of operator problem solving. The HEAP methodology has been developed to satisfy the need for objective and continuous data collection in the complex operating situation of a nuclear power plant scenario. Using a controlled experimental approach, in combination with maximised realism, has not been an obstacle for in-depth study of selected operator activities in a representative situation.

Future work in the HRP (Halden Reactor Project) will focus on how to simplify and make the methods developed more easily accessible for use in general simulator-training situations and control room validation testing. The first attempts in this direction are already taking place in the validation study of the human-machine interface for Oskarshamn Reactor unit one. Several tools have been tested in eight crews, running five scenarios at the training centre KSU in Sweden. Results from this study will be presented after a final data collection in year 2000.

- [1] Reason, J. (1990). A framework for classifying errors, in J. Rasmussen, K. Duncan, and J. Leplat (Eds.), *New Technology and Human Error*. London: John Wiley.
- [2] Hollnagel, E. (1993). The phenotype of erroneous actions. *International Journal of Man-Machine Studies*, 39, 1-32.
- [3] Skraaning, G. (1998). *The Operator Performance Assessment System (OPAS)* (HWR-538). OECD Halden Reactor Project, Halden, Norway.
- [4] Moracho, M. J. (1998). *Plant Performance Assessment System (PPAS) for crew performance evaluations. Lessons learned from an alarm study conducted in HaMMLab* (HWR-504). OECD Halden Reactor Project, Halden, Norway.
- [5] Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37, 32-64.
- [6] Hogg, D., Follesø, K., Strand-Volden, F. et al. (1995). Development of a Situation Awareness Measure to Evaluate Advanced Alarm Systems in Nuclear Power Plant Control Rooms. *Ergonomics*, 38(11), 2394-2413.
- [7] Fracker, M. L. (1991) *Measures of Situation Awareness: Review and Future Directions*. AD-A262 672, National Technical Information Service, Springfield, Virginia, 22161.
- [8] American Institute of Aeronautics and Astronautics, (1992). *Guide to human performance measurement*. (AIAA Publication No G-035-1992). Washington, D.C.
- [9] Follesø, K., Drøivoldsmo, A., Kaarstad, M., et al. (1995). *Human error – the Third Pilot Study* (HWR-430). OECD Halden Reactor Project, Halden, Norway.
- [10] Weber, M. J. (1999). *Goal Conflicts in the Process Control of a Nuclear Power Plant*. Diploma thesis submitted to the Department of Psychology, University of Bern, Switzerland.
- [11] Braarud, P. Ø. (1998). *Complexity factors and prediction of crew performance* (HWR-521). OECD Halden Reactor Project, Halden, Norway.
- [12] Braarud, P.Ø. Complexity rating of abnormal events and operator performance. *CNRA/CSNI Specialists' Meeting on Human Performance in Operational Transients*, Chattanooga, Tennessee, 1997.
- [13] Collier, S. G. (1998). *Development of a Diagnostic Complexity Questionnaire* (HWR-536). OECD Halden Reactor Project, Norway.
- [14] Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972
- [15] Ericsson, K.A. and Simon, H.A., (1980) Verbal Reports as Data. *Psychological Review*; 87: 215-251.
- [16] Pretorius, N. and Duncan, K. D. (1986). Verbal Reports in Psychological investigations: a Logical and Psychological Analysis. *Psyke & Logos*, 7, 259-287.
- [17] Kaarstad, M., Kirwan, B., Follesø, K. et al. (1994). *Human Error – the First Pilot Study* (HWR-417). OECD Halden Reactor Project, Halden, Norway.

- [18] Kaarstad, M., Follesø, K., Collier, S. et al. (1995). *Human Error – the Second Pilot Study* (HWR-421). OECD Halden Reactor Project, Halden, Norway.
- [19] Hollnagel, E., Braarud, P. Ø., Drøivoldsmo, A. et al. (1996). *The Fourth Pilot Study: Scoring and Analysis of Raw Data Types* (HWR-460). OECD Halden Reactor Project, Halden, Norway.
- [20] Braarud, P. Ø., Drøivoldsmo, A. and Hollnagel, E. (1997). *Human Error Analysis Project (HEAP) the Fourth Pilot Study: Verbal Data for Analysis of Operator Performance* (HWR-495). OECD Halden Reactor Project, Halden Norway.
- [21] Hauland, G. (1996). *Building a Methodology for Studying Cognition in Process Control: a Semantic Analysis of Visual Verbal Behaviour*. Post-Graduate Thesis in Psychology, Norwegian University of Technology and Science, Trondheim, Norway.
- [22] Kaarstad, M., Follesø, K., Collier, S. et al. (1995). *Human Error – the Second Pilot Study* (HWR-421). OECD Halden Reactor Project, Halden, Norway.
- [23] Hauland, G. & Hallbert, B. (1995). Relations between visual activity and verbalised problem solving: a preliminary study. In: Leena Norros (Ed.), *VTT symposium 158, the 5 European conference in cognitive science approaches to process control*, Espoo, Finland, pp. 99-110.
- [24] Drøivoldsmo, A., Skraaning, J., Sverrbo, M., et al. (1998). *Continuous Measures of Situation Awareness and Workload* (HWR-539). OECD Halden Reactor Project, Halden, Norway.
- [25] Skraaning, J. and Andresen, G. (1999). The Function of the Alarm System in Advanced Control Rooms: An Analysis of Operator Visual Activity during a Simulated Nuclear Power Plant Disturbance. *Paper at the Enlarged Halden Programme Group Meeting at Loen 1999*, OECD Halden Reactor Project, Halden, Norway.
- [26] Drøivoldsmo, A., Follesø, K., Holmstrøm, C. et al. (1995). *The Halden Project Workshop Meeting on Evaluation Methods and Measurements, and Data Analysis in System Test and Evaluation* (HWR-431). OECD Halden Reactor Project, Halden, Norway.
- [27] Miberg, A. B., Hollnagel, E., Skraaning, G. et al. (1999). *The impact of automation on operator performance. An explorative study*. Preliminary report. Institute of Protection and Nuclear Safety (IPSN), Paris.



# Paper F

Drøivoldsmo, A., Skraaning, G., Sverrbo, M., Dalen, J., Grimstad, T. & Andresen, G. (1998). *Continuous measures of situation awareness and workload (HWR-539)*. Halden, Norway: OECD Halden Reactor Project.





# Continuous measures of situation awareness and workload (HWR-539)

Asgeir Drøivoldsmo, Gyrd Skraaning jr., Mona Sverrbo, Jørgen Dalen, Tone Grimstad, and Gisle Andresen

## 1. OPERATOR WORKLOAD AND SITUATION AWARENESS METHODOLOGIES

### 1.1 Introduction

The studies reported here were carried out as part of the continuous activities for improving the research methodologies in HAMMLAB (Halden Man-Machine Laboratory). Experimental studies of complex problem solving in dynamic operating environments are still in their early childhood, and there is a long way to go to fulfil the needs and demands for methods and measures in this area of research. For several years, HAMMLAB experiments have been used to develop new measurement techniques, in order to fill the gap between outside expectations and the actual status of psychological methodology. The Human Error Analysis Project (HEAP) has been the central initiator in this method building activity.

The purpose of the studies described here has been to establish new and powerful techniques for operator Workload (WL) and Situation Awareness (SA) measurement. Scepticism towards prevailing methodologies, involving scenario breaks during simulator testing, has been an underlying motivation for these studies. Recurring interruptions in the scenarios represent a potential threat to the ecological validity of HAMMLAB experiments, because the questionnaires distributed during the breaks, and/or the effect of discontinuous operation, may influence human problem solving in the control room. Another reason for developing alternative SA and WL methodologies, has been the need for *continuous* measures of operator cognitive processes within critical time periods of the scenario. Critical time periods are extremely difficult to identify prior to experiments in dynamic operating situations with a multitude of solution paths. Hence, cognitive measures based upon predefined and fixed scenario breaks only account for a portion of the true variation in operator workload and situation awareness.

So far, SACRI (the Situation Awareness Control Room Inventory), and NASA-TLX (Task Load Index), have been used to evaluate SA and WL in HAMMLAB experiments (NASA, 1986; Hogg, Follesø, Torralba, & Volden, 1994). These measures presuppose scenario breaks, and are used as provisional criteria for validation of the new continuous techniques for Situation Awareness and Workload assessment.

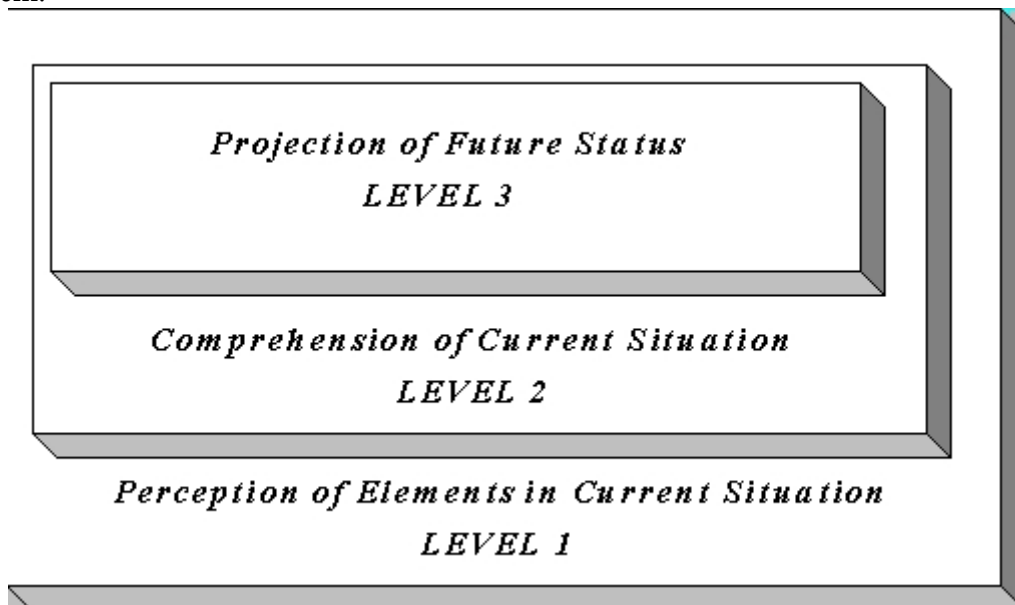
## 1.2 Situation Awareness

Situation Awareness measurement is a relatively new area in Human-Machine research and was developed within aviation, where safety researchers were searching for effective indicators of the pilots' understanding concerning the position and action of the plane. Numerous definitions of SA have been proposed, most of them pointing in the direction of 'understanding what is going on in complex systems', or 'having knowledge about dynamic situations' (Harwood, Barnett, & Wickens, 1988).

According to Endsley (1995), it is necessary to distinguish situation awareness, as a state of knowledge, from the processes of achieving that state, which is referred to as *situation assessment*. Situation awareness refers only to knowledge concerning the state of dynamic environments, and is semantically different from e.g., decision making and performance. Endsley defines SA as:

*"The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995, p36).*

In this definition, situation awareness originates from perception and active information gathering of relevant information in the environment. SA exists in the present, arising from an understanding of the past, and with predictions for how perceived and comprehended elements in the environment will have an impact on future states of the system.



*Figure 1. Endsley's Model of the Three Levels of Situation Awareness*

Endsley's model of Situation Awareness has 3 levels (Figure 1). In level 1, the operator develops an overview of the status, attributes, and dynamics of relevant information in the environment. Level 2 involves comprehension of the current situation by combining information, and forming patterns of knowledge into a holistic picture of the situation. In level 3, the operator's understanding of the situation initiates predictions about the future states of the system. The model indicates that SA is more than simply perceiving

information about the environment, but also includes comprehension of meaning, and prediction.

Endsley assumes a probabilistic relationship between situation awareness and performance. It is expected that poor performance often will occur when SA is incomplete, or inaccurate. High SA can, therefore, be viewed as a factor that will increase the probability of good performance but cannot necessarily guarantee it.

Objective assessment of SA can be made with the Situation Awareness Global Assessment Technique (SAGAT), developed by Endsley (1993). In SAGAT, operators compare the present state of the system with the past, and with normal status. They also predict how future development of the physical process will compare with the present. This information is verified against actual system states in the past, present and future, constituting a Situation Awareness estimate. SAGAT has proven useful in aviation research, and an adapted version of the system, called the Situation Awareness Control Room Inventory (SACRI), has been developed within the Halden Reactor Project (Hogg et al., 1994). SACRI requires scenario breaks in order to administer questionnaires. There is concern, however, that information held by the operators about the system will decay during these interruptions. In several studies by Endsley (1993), this was not a problem within aviation. SACRI also gives rise to other uncertainties. Questionnaire data relies heavily upon the operators' memory for time periods, and the ability to recall situations in the physical process, which may lead to confusion of system events from different time periods and consequently, unreliable measurement. The key parameters in SACRI are chosen randomly by a computer before each administration of the questionnaire. Even though the pool of parameters are representative for the scenario, the randomly selected questions presented in each break, may not be representative for the current situation. The SA questions will then be confusing, and might even lead the operator to believe that the situation is actually different from what was originally assumed. The effects of presenting questions during simulated disturbances that provide guidance, or give misleading information to the operators, should be further investigated.

### **1.3 Proposal for a new Continuous Measure of Situation Awareness**

In Endsley's definition of situation awareness (Endsley, 1995), the perception of relevant elements in the environment is the foundation for the operator's understanding of the current situation. The idea of deriving the content of thought from sensation, was first introduced by David Hume:

*"Ideas produce the images of themselves in new ideas; but as the first ideas are supposed to be derived from impressions, it still remains true, that all our simple ideas proceed either mediately or immediately, from their corresponding impressions. This is the first principle I establish in the science of human nature..." (David Hume, 1740, A Treatise of Human Nature, pp6-7).*

Hume's epistemology can be seen as the philosophical basis for the so-called *eye-mind hypothesis* (Underwood & Everatt, 1992; Hauland, 1996). According to this hypothesis, every visually perceived unit of information automatically undergoes cognitive

processing. This position is supported by the literature: Moray and Rotenberg (1989) postulate that eye fixations are closely associated with visual attention. Empirical evidence also suggests that the direction of eye movements and the direction of attention are linked (Hodgson & Müller, 1995; Shneider & Deubel, 1995).

The eye-mind hypothesis is in opposition to the well established *information-processing theory* (Best, 1992), claiming that perceived information always passes through a filter, i.e., some information is cognitively processed, while other information is not. There is reason to believe, however, that the eye-mind hypothesis can be applied in the context of simulator based experiments concerning human information-processing in complex industrial systems. Participants in such experiments are experienced professionals in their field, and any relevant information presented to them in the Human-Machine Interface is of potential significance for the operation of the system. Therefore, the operators' visual information gathering from process formats, alarm system, trend diagrams, etc., in the control room, is the main source for perception of relevant elements in the environment. This covers the first level of Endsley's model of Situation Awareness, and is easy to document from Eye-Movement tracking videos recorded in HAMMLAB. A full discussion of voluntary visual information gathering is given in Hauland (1996).

The real challenge is to measure comprehension of the current situation (Endsley's level 2), without breaking the scenarios and potentially interfering with the operators' problem solving process. Here, the eye-mind hypothesis plays an important role. A simple version of the hypothesis would claim that looking at a problem implies cognitive activities leading to an understanding of that problem. This is unreasonable, since the subject may be looking at e.g., the shape of the physical elements materialising the problem, without reference to the underlying problem syntax. A more sophisticated eye-mind hypothesis would require that the visual activity during insightful problem solving, should be a priori distinguishable from problem solving without understanding. For a puzzle problem, a certain pattern of visual fixations that will not occur spontaneously, might indicate that the problem solver is able to comprehend the problem situation. In complex systems, with thousands of components and process parameters represented in the user interface, the random probability of looking at information associated with the simulated disturbance is very low. Spending time on visual examination of such key areas in the user interface, can be seen as an indication of situation understanding (Endsley's level 2). When it comes to projection of the future status of the system (Endsley's level 3), most measures of SA are based upon the assumption that the ability to predict the future development of a system, gives little information beyond the operators understanding of the immediate process (Fracker, 1991). The relevance of Endsley's third level of SA can therefore be questioned. Eye fixations in predefined areas of the Human-Machine Interface that are important for solving problems in complex systems, can be understood as a measure of situation awareness. It is therefore proposed that the time spent by operators on examination of relevant system components during critical time periods in the scenario constitutes a new and continuous *Visual Indicator of Situation Awareness (VISA)*.

## 1.4 Workload

The function of the workload concept is to capture the amount, type and source of the load put on humans when they work. It is possible to detect and categorise a broad spectrum of factors that can lead to increased workload, e.g., the complexity of stimuli, time pressure, physical demand, etc. Consequently, it has been suggested that workload will manifest itself in different ways, resulting in a number of possible definitions, or a broad concept containing several factors.

This is reflected in a subjective rating method, called NASA-Task Load Index (NASA, 1986), which has been the standard measure of operator workload in HAMMLAB experiments for several years. NASA-TLX includes 6 sub-scales; namely mental demand, physical demand, temporal demand, performance, effort, and frustration level. Other available subjective rating methods, such as the Modified Cooper-Harper scale (Wierwille, Casali, Connor, & Rahimi, 1986), or the SWAT scale (Reid, 1981), also share the multidimensional structure of NASA-TLX, though the number and content of the sub-scales may vary. It should be noted that the factors behind the sub-scales are based upon a metaphorical conceptualisation of workload, and does not lean on a clear and straightforward definition. The NASA-TLX measure has proven to be highly reliable and applicable in a wide range of situations. However, subjective rating tests are always based upon introspection, which is a questionable methodological approach, presuming that human are fully capable of judging the amount of load put on them during a working session. Since ratings are performed in retrospect during scenario breaks, researchers have warned that response delay may lead to a loss of crucial information (Lutmer & Eggemeier, 1990). Besides, it is possible that operators put higher values on the workload scales if they are unhappy with their performance. NASA-TLX is distributed during scenario interruptions, and as for situation awareness, questions focusing on particular aspects of the operators' work might influence problem solving.

### 1.4.1 Capacity

Most theories of workload imply that the essence of mental load is that the cognitive system has a finite capacity (Reid & Nygren, 1988). Capacity is an important concept within information-processing theory (Best, 1992), and can be regarded as a crucial regulatory principle behind all theories of attention and memory. A major question concerning cognitive capacity is whether it can be described as a single resource, or several resource segments. In the first case, all kinds of mental work will compete for the same limited resource, while in the second case, there will be less total overload if the task demands do not require the same resource segment (Gopher & Douchin, 1986). The multiple resource theory has received criticism for being too complex, and it has been suggested that the single capacity model is sufficient for understanding the concept of workload (Kantowitz, 1987).

### 1.4.2 Effort

From another perspective, the meaning of workload often corresponds to the everyday concept of *effort*, which can be understood as “activities initiated to reduce the mismatch between demand and capacity” (Navon & Gopher, 1980), or “the amount of

attention and concentration directed towards a task” (Reid, Shingledecker and Eggemeier, 1981). The ability to attend/concentrate can be seen as a mental capacity, and the theoretical overlap between capacity, effort and workload is therefore evident. The reason for introducing the distinction between effort and capacity is that different measures of workload might tap into distinct aspects of the workload construct.

## **1.5 Proposal for new Continuous Measures of Workload**

### **1.5.1 A Measure of Effort**

A surprisingly direct and simple measure of workload would be to count the number of activities carried out by operators, and thereby indirectly reveal the effort put into problem solving. A classification of activity types would increase the precision of such an indicator. We have reason to believe that expert operators will have a lower activity level than operators having less experience, reflecting differences in workload. Further, it is reasonable to expect a high level of activity in complex scenarios and a lower level of activity in simple scenarios, also representing a difference in workload. These expectations are based on the assumption that the various levels of operator activity is an indicator of the demand put on the information-processing system and the corresponding effort initiated to meet these demands. It is of particular interest to note that NASA-TLX can be interpreted as effort-oriented, asking questions about the physical and mental demand of tasks, and the degree of experienced effort and frustration during problem solving. Intuitively at least, such questions can be associated with the actual number of tasks performed by the operator during scenarios. Therefore, the accumulated *Activity Level* for different types of actions in the Human-Machine Interface is proposed as a new continuous measure of operator effort.

### **1.5.2 A Measure of Capacity**

It is known that the eye can recognise simple information, e.g., a letter, within 50ms (Sperling, 1960). On a subconscious level, information can be registered at even shorter times. More complex stimuli presented in parallel with higher cognitive processing require fixation times from 100-200ms as a minimum (Optician, 1985; Carl & Gellman, 1987). A group of fixations representing the minimum gaze necessary to encode a single information unit during problem solving can be referred to as a "dwell" (Hauland 1996). Assuming a single capacity model, it seems logical that dwell times will be extended as a function of mental demands required by the task. This was supported by Svensson, Angelborg-Thanderz, Sjöberg, & Olsson (1997), finding that fixation times increased as a linear function of task difficulty. Durso et. al. (1995) were able to show that novice chess players generally had a longer fixation duration than experts and intermediate players. The psychological hypothesis behind the idea of extended dwell-time for demanding cognitive activities, is that information encoding will suffer during periods of increased ‘high level cognitive processing’, because the total capacity of working memory is limited. *Extended Dwell-Time (EDT)* is therefore proposed as a new continuous measure of capacity.

## 2. GENERAL METHOD

The new measures of situation awareness and workload have been tested and evaluated in two experiments: the 1996 CASH/ NRC alarm system experiment (O'Hara et al., 1997), and the 1997 Human Error Analysis Project & Human-Centred Automation (HEAP/HCA) experiment. Both studies were carried out under controlled conditions in the HAMMLAB simulator. In the CASH alarm experiment, 6 crews of 2 professional operators from the Loviisa power plant in Finland were tested in 16 scenarios, covering a range of complex and realistic operating situations. In the HEAP/HCA experiment, 15 operators from the Halden Boiling Water Reactor participated. They were tested individually in 6 different scenarios.

*Randomised block (factorial) design* with operator or crew as the blocking variable, was applied in both experiments (Kirk, 1995), producing a high level of statistical power. Since order effects are a serious threat to the validity of within-subject designs, the presentation sequence of scenarios and experimental conditions was counterbalanced (Jones, 1995). The main dependent variables in both experiments were operator performance, plant performance, situation awareness, and workload. Corresponding measurement techniques were the Operator Performance Assessment System (Skraaning, 1998), the Plant Performance Assessment System (Moracho, 1998), the Situation Awareness Control Room Inventory (Hogg et al., 1994), and the NASA-Task Load Index (NASA, 1986). The new continuous SA and WL measures were calculated for a selection of operators in the CASH alarm experiment, and all operators for 2 of the 6 scenarios in the HEAP/HCA experiment.

### 2.1 Selection of Operators

In the CASH alarm study, Eye Movement Tracking (EMT) was recorded for only 1 operator in every second scenario. A set of operators, therefore, had to be selected in order to develop and evaluate the new EMT-based measures of SA and WL (VISA and EDT). In addition to the selection imposed by the eye-tracking data, operators and scenarios in the CASH alarm experiment were chosen according to the *Active Operator Principle (AOP)*. This principle states, that in a scenario where two or more operators are working together to solve problems, the operator with the highest load and most problem-relevant tasks, should represent the crew's problem solving in the scenario, e.g., if the critical action in the scenario is to stop a leakage in the primary circuit, the Reactor Operator (RO) is chosen as "the active operator", since his/her problem solving and interaction with the simulator interface will be most representative of crew's performance.

To avoid work on the primary side being overly influenced by the Turbine Operator (TO), crews where the RO held a Shift Supervisor licence were selected. This was to ensure similar communication and co-operation patterns in the crews. Crew 1 and 6 were not included in the study because the TOs were trained as Shift Supervisors in these crews. The study made use of all the primary side scenarios where the ROs' eye movements were recorded, and all the secondary side scenarios where the TOs' eye movements were recorded. Two periods were scored in each scenario, but due to technical problems, one period was left out for one of the operators.

In the HEAP/HCA experiment, half of the cases in 2 scenarios had breaks with collection of SACRI and NASA-TLX data. All 15 operators were analysed for the scenario in which they were tested with scenario interruptions. As was the case for the CASH alarm scenarios, each scenario was divided into two periods.

## **2.2 Scenarios**

To maximise external validity, it was important to evaluate the new SA and WL measures in various operating situations. The following scenarios were used:

For the CASH alarm study, subject matter experts with extensive Loviisa experience developed 16 scenarios. 14 of these scenarios were selected to test the new continuous SA and WL measures. The scenarios used were: (1) failure of generator to trip, (2) reactor scram, (3) reactor coolant pump LOCA, (4) turbine trip, (5) condensate valve coupling failure, (6) superheater malfunction and TC-controller failure, (7) air leakage in valve between condenser and ejector, (8) turbine overspeed, (9) oil in compressed air (pneumatic) system, (10) loss of main transformer-extreme weather-snow, (11) instrumentation line leakage, (12) small LOCA, (13) steam generator tube rupture, and (14) cycling of MSIV-secondary pressure transient-main steam line break.

For the HCA/HEAP experiment, a set of 6 scenarios was developed. As stated above, two of these scenarios were selected. They were (1) leakage in the purification system and (2) leakage in the sealing water system.

## **2.3 Apparatus**

Data from the simulator were recorded automatically in the experimenters log. The experimenters log includes a record of the alarm system and all the operators' interactions with the simulator interface.

Eye movements were recorded with an ASL 4000 system, using a direct scene camera, and logging data at a frequency of 50Hz. The EyeCon system and the eye-tracker scene camera video with the cross hair were used for the analysis. The apparatus used for scoring the VISA measure is described in Appendix I. The eye-tracker data log was used for scoring of the EDT workload measure.

All log-transformations and statistical analysis were performed in Statistica (Statsoft, 1997), on Excel worksheets (Microsoft, 1996), or with special scripts written in the Pearl programming language.

## **2.4 Procedure for Data Scoring and Analysis**

SACRI and NASA-TLX data from both experiments were scored according to the procedure used in Hogg et al. (1994) and Hallbert et al. (1996). The procedure for scoring the VISA measure was according to the same general principles in both experiments, but the type of scoring elements varied slightly (see section VISA Scoring Technique). The method for scoring EDT and Activity Level was identical in the experiments.



## 2.5 VISA Scoring Technique

The VISA scoring technique consisted of three main steps.

1. A process expert defined Areas Of Interest (AOIs) and critical time-windows for each scenario. The AOIs were parts of the simulator interface, necessary for information gathering and problem solving in the scenario. In the CASH alarm study, AOIs were defined as sub-systems. This is exemplified as area in Figure 2.
2. In the HCA/HEAP experiment, AOIs were scored as process format, area, or components, as shown in Figure 2.
3. The AOI scoring technique was used for all VISA-scoring (Hauland & Hallbert 1995), and the areas of interest selected are listed in Appendix I.

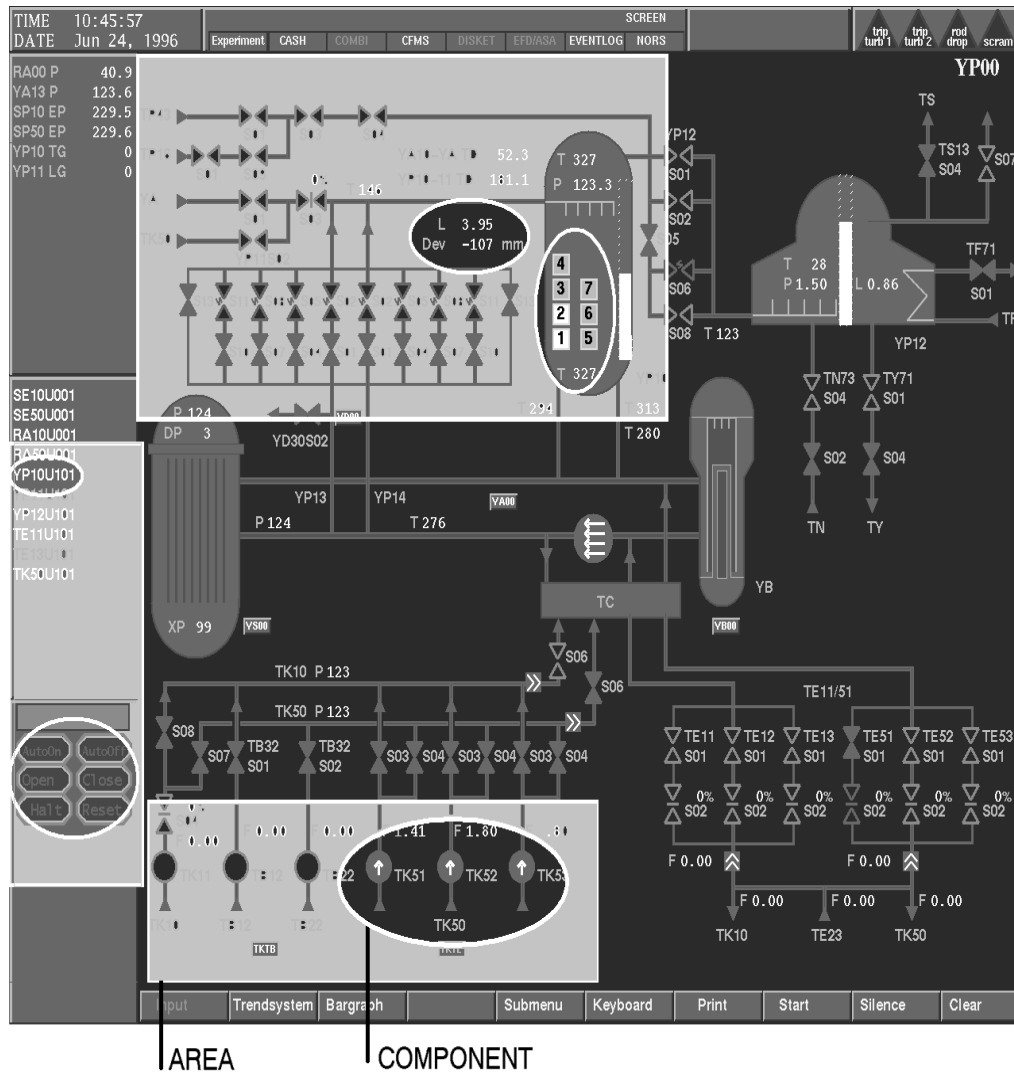


Figure 2. Example of Areas Of Interest on Process Format, Area and Component Level.

4. Scoring of dwell length in EyeCon was performed using the following procedure. When the operator looked at an indicator for less than one second, it was scored as *short* dwell, i.e., the length of a dwell was set to the mean between a minimum dwell (0.2 seconds) and a maximum dwell (1.0 second). A short dwell is therefore counted as 0,6 seconds. If the time spent looking at the area was longer than one second, the following procedure was followed: the scorer first pushed the button for the relevant area when the operator started looking at it. When the operator stopped looking at this area, the scorer pushed a button defined as LONG. This procedure gave a time tag in the log for when the dwell started, and a time tag for when the dwell ended. By using these two time-tags, *long* dwell time was measured.
5. The EyeCon data log was transformed to scores and the total time spent looking at the AOIs was calculated and copied into a final result matrix. The scenarios varied in both total length and in period length. Two periods were used from each scenario. The difference in time is accounted for in the analyses, as the time the operator monitors relevant information in the process, per time unit.

In scenarios with more than 10 AOIs the scorer had to look through the same video twice, scoring half of the AOIs in each round.

## 2.6 Workload Scoring Technique

Workload was scored by two measures, EDT and Activity Level. EDT was calculated in the Eyetrack II program (RISØ National Laboratory) using visual angle coefficients of 5 and 7, and a minimum dwell time of 150 ms. A detailed description of this procedure can be found in Appendix II. The Activity Level was measured through semi-automatic counting of operator actions for different types of activities in the experimenters log.

## 3. EXPERIMENT 1-1. SA (CASH ALARM EXPERIMENT)

### 3.1 Method

The CASH alarm study scenarios were used in this experiment. Operators were selected by the Active Operator Principle described in chapter Selection of Operators.

The VISA scoring in this experiment had two different categories of AOIs. The first category was *general* AOIs, representing plant information, and in accordance with the critical parameters defined in the Plant Performance Assessment System (Moracho, 1998). The other category was *specific* AOIs based upon important activities defined in the Operator Performance Assessment System (Skraaning, 1998). The resolution of the eye-tracker scene camera was not high enough to read small letters in the process displays. Therefore, trend curves were scored using information from the experimenters log in combination with the eye-tracker. The experimenters log contained information about which trends the operator had open on a given monitor at a certain time.

The total time spent looking at each indicator was computed. This left two parameters for each of the two periods in each scenario:

1. Specific time: Percentage of time spent monitoring specific AOIs per total time of the period [0,100].
2. General time: Percentage of time spent monitoring general AOIs per total time of the period [0,100].

### 3.2 Results

The results showed a significant negative correlation between SACRI A' (the main SA score in SACRI) and VISA specific time (Table I and Figure 3). SACRI A' refers to the test subject's accuracy in detecting drifts in the NORS process monitoring task. A high A' score indicates that the test subject has correctly diagnosed the process. The negative correlation between the SA measures, means that higher Situation Awareness corresponds to shorter total monitoring of specific VISA indicators. This is in accordance with theoretical and empirical findings within cognitive psychology. Preconceptions about the characteristics of the information sought for, will influence the speed and accuracy of information gathering. If the information is in agreement with the subject's expectation, this will facilitate the speed of information gathering, and vice versa. However, if the information is deviating from the subject's expectation this will inhibit the speed of information gathering. For example, the classical method of *priming*, produces shorter recognition times for expected stimuli, while unexpected stimuli inhibits the reaction time (Meyer & Schvaneveldt 1971, Neely, 1991).

Table I.

	General Time	Specific Time
Total time SACRI A'	n.s.	r= -0,52, n=31, p<0.01
First period SACRI A'	n.s.	r= -0.86, n=16, p<0.01
Second period SACRI A'	n.s.	n.s.

A significant effect of period ( $F(1,15)=5,77; p < 0,05$ ) was found for VISA specific time. No such effect was observed for the SACRI A' measure. Monitoring of general SA indicators was not associated with SA, as measured by SACRI. Neither was the number of times the operator looked at AOIs. A possible explanation for this, can be that the analysis did not take into account the number of AOIs actually looked at, i.e., if operators with low average specific time collected information from a higher number of relevant AOIs, this would give a higher number of hits, but the specific time would still be low. Another explanation would be that the frequency of AOI-accesses and the time spent in an AOI, to a certain extent would be antagonistic properties. If the operator spent a long time looking at an AOI, this would give him less time to check other relevant information in the interface.

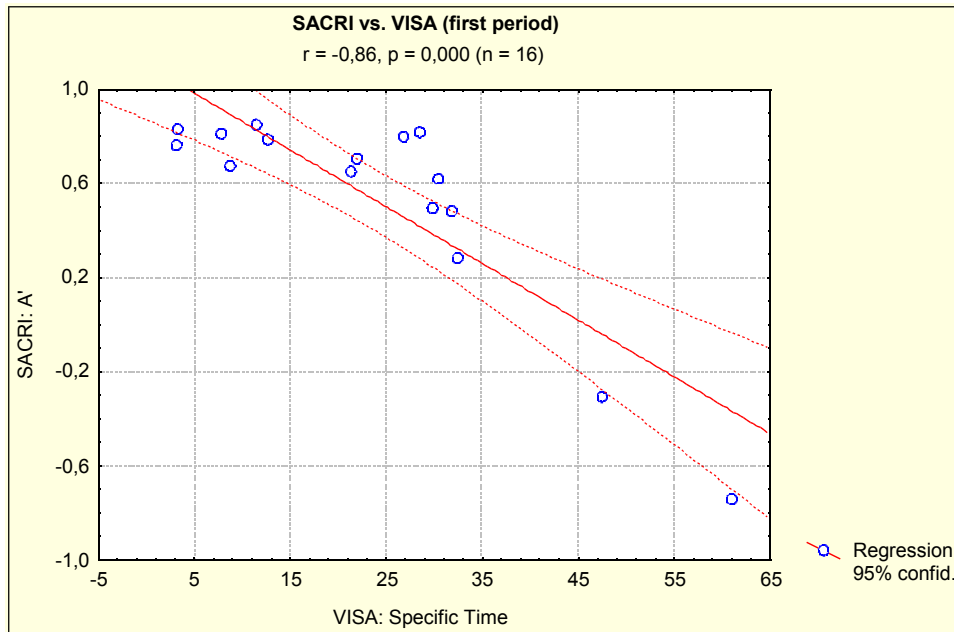


Figure 3. Correlation between SACRI and VISA in the First Scenario Period.

### 3.3 Discussion

There is reason to believe that the general indicators in VISA did not produce significant results, because experienced operators are able to understand the state of the plant indirectly, without looking directly at the main plant parameters, or by looking at the trends, which could be exposed inaccuracy in the scoring.

It can be concluded that the less time the operators look at specific indicators, the higher is their SA, according to SACRI. This effect is probably limited to the phase of scenarios when operators' are most active in their problem solving.

## 4. EXPERIMENT 2-1. SA (HEAP/HCA EXPERIMENT)

It is reasonable to expect a significant difference in performance on interface operations between operators in the first and second experiment. Using 49 crews formed by experienced operators, Fujita and Toquam (1997) found as much as 20% of the performance score variance between operators, was explained by job and training experience. The training and experience difference between the two experiments in this study was considerable. Before the HEAP/HCA experiment, Halden operators had only finished the first phase of their NORS training, and in addition they had a delay between training and the experimental tests of about one month. On the other hand, the Loviisa operators were quite familiar with the NORS simulator from their own plant, and they had no delay between two days of interface training and the experimental tests. Following these facts, the hypothesis for the HEAP/HCA experiment was that the

connection between SACRI and VISA could be quite different from the first experiment.

In literature, expectancy has been used to account for priming effects (Becker, 1980, 1985, Posner and Snyder, 1975). Expectancy theories account for the standard priming effect by assuming that subjects use the prime to generate an expectancy set that consists of potential targets that are related to the prime. Following this, the operators should spend less time looking at critical process parameters when they understand the situation and these parameters are in accordance with their expectancies. As reported above, this was found in the CASH alarm SA-experiment. Dependent on operators experience, it is reasonable to assume that their expectancies will vary from no expectancy set (no situation awareness), via varying degrees of unrealistic expectancies (erroneous hypothesis about the problem) with some realistic and some unrealistic expectancies (both correct, and erroneous hypothesis), to a high number of realistic expectancies. It is also realistic to assume that expectancy will affect the operators' time for information gathering. Incorrect expectancies will lead to prolonged time for information gathering. Inhibition effects in lexical decision tasks are typically found for unrelated prime and target (Neely, 1977, Favereau & Segalowitz, 1983, den Heyer et al., 1985, and Smith et al., 1987), i.e., when the prime gives the subject expectations about something which is to happen, and something else happens instead.

In Posner and Snyder's (1975) theory, expectancy is: (1) under a person's strategic control and cannot occur without a person's intention or awareness, (2) relatively slow acting in that it takes time to generate the expectancy set from the prime, (3) facilitates the processing of expected targets, whether they are related or unrelated to the prime, and (4) inhibits the recognition of unexpected targets, whether they are related or unrelated to the prime, from Neely (1991).

Before the HEAP/HCA experiment, a curve for visual information gathering from critical AOIs as a function of the operators' experience level, like the one shown in Figure 4, was predicted from theory. The general assumption was that the relationship between SACRI and VISA would be moderated by the operator level of experience. Since the SACRI A' score will increase gradually with the operators' experience and training level, the association between the SA measures was expected to be positive for low levels of experience, and negative for higher levels of experience. Because of uncertainty about the Halden operators' level of training, no attempt was made to predict the exact relationship between VISA and SACRI in the HEAP/HCA experiment.

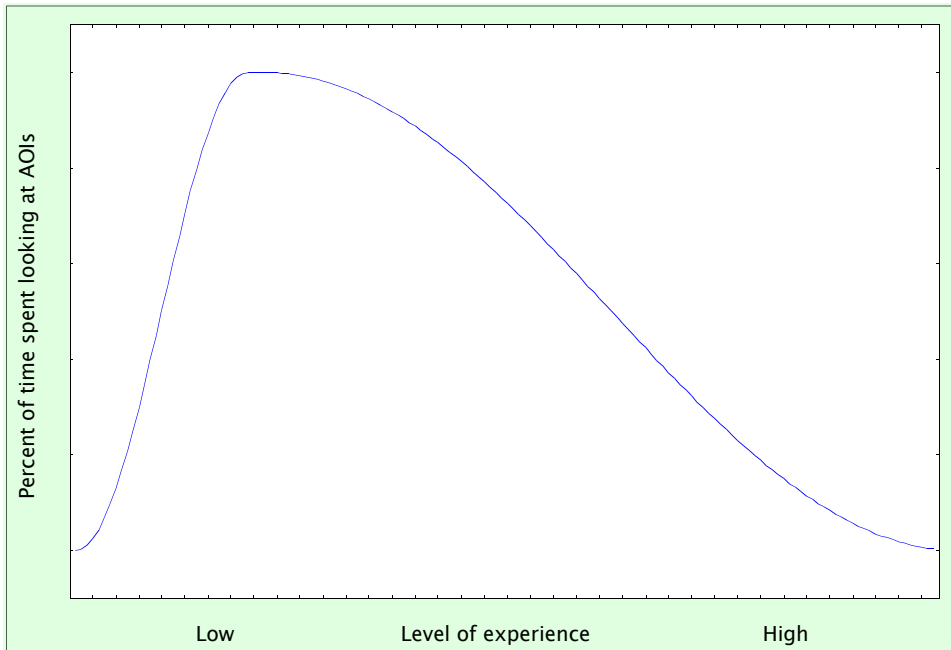


Figure 4. Hypothesised Curve Describing the Connection between VISA and Level of Experience.

This study used two HEAP/HCA scenarios. All 15 operators participating in the study were scored according to the method described in section Procedure for Data Scoring and Analysis. The VISA scoring in this experiment contained three different levels of AOIs as described in Figure 2.

The total time spent looking at each indicator was computed for each level. This left three parameters for each of the two periods in each scenario:

1. Format time: Percentage of time spent monitoring process formats and overview displays with indicators per total time of the period [0,100].
2. Area time: Percentage of time spent monitoring areas in the interface with indicators per total time of the period [0,100].
3. Component time: Percentage of time spent monitoring components in the interface with indicators per total time of the period [0,100].

## 4.1 Results

As shown in Table II., the correlation between SACRI and VISA in this study was positive, i.e. more time spent looking at the critical parameters gave higher SA measured by SACRI.

Table II.

Variable	VISA, format time	VISA, area time	VISA, component time
SACRI A'	0.53, p< 0.05	0.37, n.s.	0.60, p< 0.05

The positive correlation can be explained by the leftmost part of the curve shown in Figure 4. Assuming that operators process more in critical AOIs and had more expectations about what happened, and that operator skill and experience with the simulator lead to higher SA, the operators with the strongest expectations suffered most from insufficient training and used more time to correct their expectations.

## **4.2 Discussion**

The operators in this study were moderately trained in detection and localisation of problems, but not trained in handling operations to stabilise the plant. With a minimum of training in the interface, the maximum level of performance in this experiment was to identify the right indicators and use them for diagnosing the situation. Trying to correct the plant situation, the operators made many mistakes and they were unsure about the effect of their operations. Execution of control operations would therefore initiate more expectations with a consequent need to confirm them. This can be seen through the operators' constant need for confirming the essential parameters to keep control of the situation and maintain SA. This effect has also been shown in experiments comparing novice and skilled readers. For example in reading of music, the skilled musician will have a smoother reading of the notes, while the beginners will dwell with their mistakes. "The novice's eyes will not only drift back a beat or two to notes they have misplayed, but jump back several bars to check the key signature, which is inscribed at the beginning of each line of music." (Furneaux, 1997). Transferred to nuclear power plant control, this means that the less trained operator will check several of the critical parameters more frequently, since he is unsure about causes and effects.

## **5. EXPERIMENT 1-2. WL (CASH ALARM EXPERIMENT)**

### **5.1 Method**

Workload was measured by NASA-TLX and the two proposed measures EDT, and activity level. The EDT measure was the operators mean dwell-time in a scenario period. The dwell-time was calculated from the eye-tracker data log with a minimum dwell time of 150 ms. and a visual angle parameter of five (see Appendix II for a detailed description of the EDT scoring).

The scoring of activity level involved a systematic counting of responses from the experiment log. Different actions from the experiment log were grouped into five variables: navigation, information gathering, operations, alarm interactions and format shifts. This categorisation followed practical and theoretical assumptions. Navigation included all actions indicating that the operator was actively searching, and answering demands from the system. Information gathering referred to the operator asking for information about a component or information on automatics. Operations where a group of actions, like opening or closing valves or switching from automatic to manual control. Format shifts were connected to information gathering. Appendix III shows the actions included in the five groups.

Data were transformed to z-scores based on the duration of scenario periods. This gave the activity level for each operator in different conditions. Multiple Regression analysis was used to find which of the groups contributed to explain the variation in workload. This left us with three significant groups: Information gathering, operations, and alarm interaction. These groups were added and called "Activity Level", and were used further in correlation and ANOVA analysis.

The same scenarios and subjects were used as in the SA CASH alarm study. The active operator principle was applied, and workload measures were scored for both periods in each scenario.

## 5.2 Results

### 5.2.1 Measuring Capacity

The CASH alarm study showed an overall correlation between EDT and NASA-TLX ( $r=0.46$ ,  $n=35$ ,  $p< 0.01$ ). For the first and second period, the correlation was  $r= 0.57$  and  $r= 0.52$  ( $p< 0.05$ ) respectively. As predicted, these results indicate longer dwell time as a function of higher workload.

Comparing the capacity workload measures with SACRI, Table III. shows that higher SA leads to reduced workload (or vice versa) measured for the first scenario period. For the second scenario period no relationship was found. This can be compared with findings in the CASH alarm SA study, where VISA specific time also correlated with NASA-TLX ( $r= 0.63$ ,  $n=31$ ,  $P< 0.01$ ) for both periods, and ( $r= 0.71$ ,  $n=16$ ,  $p< 0.01$ ) for the first period only. (Note, the direction of the VISA measure is opposite of SACRI.)

Table III.

	SACRI whole scenario (n=35)	SACRI first period (n=18)	SACRI second period (n=17)
NASA-TLX	n.s.	$r= -0.45$ , $p=0.06$ (n.s.)	n.s.
EDT	$r= -0.44$ , $p< 0.01$	$r= -0.60$ , $p< 0.01$	n.s.

Relating these results to the curve in Figure 4, it could be reasoned that the correlation between workload and experience is negative, which is in accordance with known theory.

### 5.2.2 Measuring Effort

As with NASA-TLX, Activity Level measurement was also sensitive to periods in the scenarios. Using ANOVA analysis a main effect for Activity Level was found ( $F(1,8) = 6,97$ ;  $p< 0,05$ ) between the first and second period in the scenarios.

Differences in the operators activities varied the same way as in NASA-TLX, the relationship is shown in Figure 5. Higher activity level leads to higher workload.

As was mentioned above, for a workload measurement technique to be recommended, it should have certain properties. Primary among them are *sensitivity*, which is the degree



to which a given workload technique can distinguish differences in levels of load imposed on an operator. Activity Level has in this sense showed sensitive.

*Intrusion* is another criteria, which is an undesirable property in which introduction of the workload measuring technique causes a change in operator-system performance. This criteria is addressed in Activity Level measures, as scenarios can be run without any break or interruptions. A third criterion considers the degree of *diagnosticity*, i.e., the ability to discern the type or cause of workload, or the ability to attribute it to an aspect of the operators task. An obvious advantage of the Activity Level measure is the opportunity to measure any time window during a scenario, related directly to the operators problem solving. The last aspect concerns *transferability*, which is the capability of a technique to be used in various applications. Usually the primary task will vary from application to application, but through its simple data logging procedure, Activity Level measurement is promising for a variety of different settings.

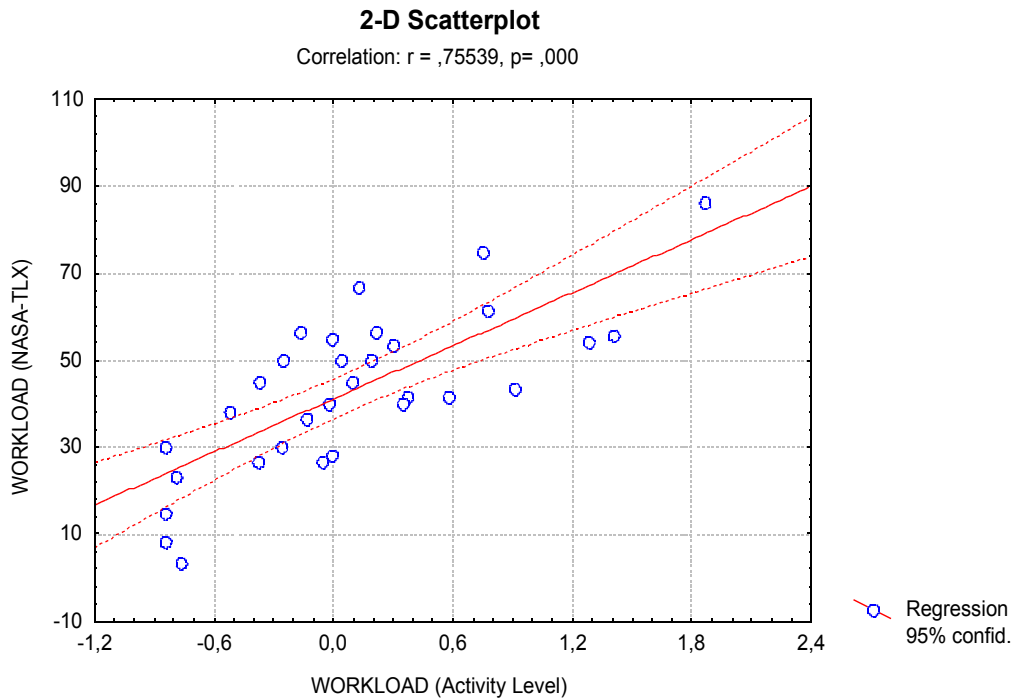


Figure 5. Correlation between Activity Level and NASA-TLX

Non-intrusive workload techniques that have a global sensitivity are of utmost importance (Wierwille & Eggemeier, 1993). According to the criteria above, analysis of operator activity level has a promising potential. This method should be further tested in new experiments with varying conditions.

## **6. EXPERIMENT 2-2. WL (HEAP/HCA EXPERIMENT)**

### **6.1 Method**

Workload was measured by NASA-TLX, EDT and Activity Level. The same scoring procedure was used as in the preceding CASH alarm workload experiment. Minimum dwell time was set to 150 ms. and the visual angle parameter was 7. 15 operators and 2 scenarios, as described in Chapter 2, were used for all measures.

### **6.2 Results**

Correlation was used to evaluate the relationship between EDT and NASA-TLX. No significant correlation was found. In this study the NASA-TLX showed no explanatory power at all. An important property of this method is that the weighting is decided by the subject themselves. In other words, the relative contribution of the sub-scales to the total workload is based on the operators' own experience, not on a priori understanding and definition of the workload concept. Operators with less experience can suffer from several influencing factors, such as training effects, interface problems and so forth. The missing correlation between workload and task complexity can be regarded as an indication of missing construct validity in this study. Given this situation the subjective rating of NASA-TLX could suffer from undesired biases caused by a number of different sources of noise.

Comparing EDT with the VISA measure, a positive correlation was found on the VISA component time ( $r=0.45$ ,  $n=30$ ,  $p<0.05$ ). Significance was obtained only for the second period ( $r=0.59$ ,  $n=15$ ,  $p<0.05$ ). The fact that the correlation was positive indicates that for less experienced operators the relationship between VISA and EDT was essentially the same as for more experienced operators.

Activity level was also measured in this experiment, but due to the uncertainty in the NASA-TLX measure, analyses of this measure have been regarded less interesting. Though, Activity Level and EDT correlated negatively ( $r=-0.54$ ,  $n=15$ ,  $p<0.05$ ) for the first period of the scenarios, indicating that the same pattern with differences between the two pools of operators was present. An effect of period was also obtained with higher activity level in the second period ( $F(1,26)=4.38$ ,  $p<0.05$ ).

## **7. CONCLUSION**

Despite the preliminary status of these studies, it can be concluded that the VISA measure is promising as an indicator of SA. However, there is a need for development of the theoretical foundation and the practical implementation of the measure. The effort for scoring VISA has shown to be about the same as for scoring SACRI. One major advantage of the VISA measure is the opportunity to define critical time windows for the scoring post hoc, dependent on the critical events and the relevant process parameters in the scenario. In a complex situation like a nuclear power plant, the timing

of scenario breaks has shown to be difficult and continuous measures of workload and situation awareness will be most valuable.

The goal should be a robust measurement system which can be used as a tool in a number of different settings. The observation of opposite polarity for the SA and WL measures on different operator experience levels, could be investigated in longitudinal studies following operators during training. If these effects are shown to be strong, the measures would be valuable contributions to mapping of training level and experience.

Future development of the VISA measure should consider several additional factors. A strengthening of level three in Endsley's definition would be to combine a verbal component with the eye movements. A combination of the current VISA technique with the operators verbalisation about the future status of system should be tested.

To strengthen the construct and content validity of VISA and EDT, techniques should be developed to connect the measures more closely with the variable log and the actual fluctuation of the variables defined as AOIs. There will also be a need for reliability testing of the new measures.

## 8. REFERENCES

- Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis in semantic strategies. *Memory and Cognition*, 8, 493-512.
- Becker, C. A. (1985). What do we really know about semantic context effects during reading? In D. Besner, T. G. Waller, & E. M. MacKinnon (Eds.), *Reading research: Advances in theory and practice*, 5 (pp. 125-166). Toronto: Academic Press.
- Best, J. B. (1992). *Cognitive Psychology*. St. Paul: West Publishing Company.
- Carl, J. R., & Gellman, R. (1987). Human Smooth Pursuit: Stimulus-Dependent Responses. *Journal of Neurophysiology*, 55, 1444-1463.
- Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, M., Nikolic, D., Moertl, P. M., Ohrt, D., and Manning, C. A. (1995). Expertise and Chess: A Pilot Study Comparing Situation Awareness Methodologies. In Garland, D. P. & Endsley, M. R. (Eds.), *Experimental Analysis and Measurement of Situation Awareness* (pp. 295-303). Embry-Riddle Aeeronotical University Press.
- Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems, *Human Factors*, 37, 32 - 64.
- Endsley, M. R. (1993). Situation Awareness in Dynamic Human Decision Making: Theory. *Proceedings of the 1st International Conference on Situational Awareness in Complex Systems*, Orlando, February 1993.
- Favreau, M. & Segalowitz, N. S. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory and Cognition*, 11, 565-574.
- Furieux, S. (1997). Professionals know the score. S. Furieux interviewed by M. Hamer. *Scientific American*, 15 February, 1997.
- Fracker, M. L. (1991). *Measures of situation awareness: Review and future directions*. Wright-Patterson AFB OH: Armstrong Laboratory, Crew Systems Directorate.
- Fujita Y., Toquam, J.L. (1997). Assessing nuclear power plant operator variability. In N. Stanton (Eds). *Human Factors in Nuclear safety*. University of Southampton. Taylor & Francis.
- Gopher, D., & Donchin, E. (1986). Workload-An examination of the concept. In K. R. Boff, L. Kaufman and J. P. Thomas (Eds.), *Handbook of perception and performance: Volume II. Cognitive processes and performance* (pp. 41-1-41-49). New York: Wiley.
- Harwood, K., Barnett, B., & Wickens, C. (1988). Situational awareness: A conceptual and methodological framework. Proceedings of the symposium Psychological in the Department of Defence (pp. 316 - 320) Colorado Springs, CO: US Air Force Academy.

- Hallbert, B., Sebok, A., Persensky, J., Morisseau, D. (1996). *Results of the Study of Control Room Crew Staffing for Advanced Passive Reactor Plant*. OECD Halden Reactor Project, HPR -348, Halden, Norway.
- Hauland, G. & Hallbert, B. (1995). Relations between visual activity and verbalised problem solving: a preliminary study. In: Leena Norros (Ed.), *VTT symposium 158, the 5<sup>th</sup> European conference in cognitive science approaches to process control*, Espoo, Finland, pp. 99-110.
- Hauland, G. (1996). *Building a methodology for studying cognition in process control: a Semantic Analysis of Visual and Verbal Behaviour*. Unpublished master's thesis, Norwegian University of Science and Technology (NTNU), Department of Psychology, Trondheim.
- den Heyer, K., Briand, K., & Smith, L. (1985). Automatic and strategic factors in semantic priming: An examination of Becker's model. *Memory and Cognition*, 13, 228-232.
- Hodgson, T. L., & Müller H.J. (1995). Evidence relating to premotor theories of visuospatial attention. In J.M. Findlay et al. (Eds.): *Eye Movement Research*. Elsevier Science B.V.
- Hogg, D. N., Follesø, K., Torralba, B., Volden, F. S. (1994). *Measurements of the Operator's Situation Awareness for Use Within Process Control Research: Four Methodological Studies*. OECD Halden Reactor Project, HWR-377, Halden, Norway.
- Hume, D. (trans 1978). *A Treatise of Human Nature*, Analytical Index by L.A. Selby-Bigge, text revised and notes by P.H. Nidditch. Oxford: Oxford University Press.
- Jones, J. L. (1995). *Understanding psychological science*. New York, NY: Harper Collins.
- Kantowitz, B. H. (1987). Mental Workload. In P.A. Hancock (Ed.), *Human Factors Psychology*. Amsterdam: North Holland.
- Kirk, R. E. (1995). *Experimental design: procedures for the behavioural sciences*, 3<sup>rd</sup> ed. Pacific Grove, CA: Brooks Cole.
- Lutmer, P. A., & Eggemeier, F. T. (1990). *Effect of intervening task performance on subjective ratings of mental workload*. Presented at the Fifth Mid-Central Ergonomics Meeting, Dayton, OH.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*. 90, 227-234.
- Microsoft Excel 97. Copyright 1985 - 1996. Microsoft Corporation.
- Moracho, M. J. (1998). *Plant Performance Assessment System (PPAS) for crew performance evaluations. Lessons learned from an alarm study conducted in HAMMLAB*. OECD Halden Reactor Project, HWR-504, Halden, Norway.
- Moray, N., & Rotenberg, I. (1989). Fault management in process control: eye movements and action. *Ergonomics*, 32 ( 11). 1319-1342.

- NASA-Ames Research Centre, Human Performance Group (1986). *Collecting NASA workload ratings: A paper-and-pencil package (Version 1.1)*. Moffet Field, CA: NASA-Ames Research Center.
- Naveon, D., & Gopher, D. (1980). Task difficulty, resources, and dual task performance. In R. S. Nickerson (Ed.), *Attention and Performance VIII* (pp. 297-315). Hillsdale, NJ: Erlbaum.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Erlbaum.
- Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Journal of Experimental Psychology: General*, 106, 226-254.
- O'Hara, J., Brown, W. S., Hallbert, B., Skraaning, G., Persensky, J. J., & Wachtel, J. (1997). *The effects of alarm processing and display on operator and plant performance*. Preliminary report, US Nuclear Regulatory Commission, Washington, D.C.
- Optican, L. M. (1985): Adaptive Properties of the Saccadic System. In: A. Berthoz & M. Jones (Eds.), *Adaptive Mechanisms in Gaze Control: Facts and Theories*. Elsevier Science Publishers, New York, pp. 71-79.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, NJ: Lawrence Erlbaum. Associates.
- Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (Eds.), *Human mental Workload* (pp. 185-218). Amsterdam: Elsevier.
- Reid, G. B., Shingledecker, C. A., and Eggemeier, F. T. (1981). Application of conjoint measurement to workload scale development. In *Proceedings of the Human Factors Society 25<sup>th</sup> Annual Meeting* (pp. 522-525). Santa Monica, CA: Human Factors Society.
- Shneider, W. X., & Deubel, H. (1995). Visual Attention and saccadic eye movements: Evidence for obligatory and selective spatial coupling. In Findlay et al. (Eds.): *Eye Movement Research*. Elsevier Science B.V.
- Skråning, G. (1998). The Operator Performance Assessment System (OPAS). OECD Halden Reactor Project, HWR-538, Halden, Norway.
- Smith, M.C., Briand, K., Klein, R.M. & Den Heyer, K. (1987). On the generality of Becker's Verification Model. *Canadian Journal of Psychology*, 41, 379-386.
- Sperling, G. A. (1960). The information available in brief visual presentation. *Psychological Monographs*, 74, Whole No. 498.
- StatSoft, Inc. (1997). *STATISTICA for Windows* [Computer program manual]. Tulsa, OK.

- Svensson, E., Angelborg-Thanderz, Sjöberg, L., Olsson, S., (1997). Information complexity - mental workload and performance in combat aircraft, *Ergonomics*, vol 40.(pp. 362-380).
- Underwood, G., & Everatt, J. (1992). The Role of Eye Movements in Reading: some limitations of the eye-mind assumption. In E.Chekaluk and K.R. Llewellyn (Eds.), *The Role of Eye movements in Perceptual Processes*, Elsevier Science Publishers B.V., Advances in Psychology. Amsterdam, Vol. 88, pp. 111 – 169.
- Wierwille, W. W., Casali, J. G., Connor, S. A., & Rahimi, M. (1986). Evaluation of the sensitivity and intrusion of mental workload estimation techniques. *Advances in man-machine systems research* (Vol. 2, pp. 51-127). Greenwich: JAI Press, Inc.
- Wierwille & Eggemeier (1993). Recommendations for Mental Workload, Measurement in a Test and Evaluation Environment. *Human Factors*, 35 (2), 263-281.

## **9. APPENDIX I**

Description of the procedure for scoring VISA.

The scoring of VISA was done with the EyeCon system. In this system the computer log is synchronised to the video players time code, which allows systematic registration of areas of interest (AOI), supplemented with time tagging. The registration was done by pressing a touch sensitive tablet connected to the computer running EyeCon.

Before the registration started, a template was made, defining which areas on the tablet should represent which AOI. The CASH alarm study used templates on the area level, while the HEAP/HCA study used three different levels of AOIs: the format level, representing the whole area in the specific display-format; the component level, representing a specific component in the display-format; and the area level, the area surrounding the component level. The AOIs were areas containing critical information during the scenarios, chosen by a process expert. AOIs for the two experiments are shown in the tables below.



*Areas of Interest used during VISA scoring, HEAP/HCA experiment*

<b>Scenario</b>	<b>Area of interest</b>	<b>Area of interest level</b>
<b>3</b>	TC 00 TC 00 TC 00 ALARM OVERVIEW OVERVIEW OVERVIEW POWER POWER POWER	format level area level component level format level format level area level component level format level area level component level
<b>4</b>	YD 00 YD 00 YD 00 ALARM OVERVIEW OVERVIEW OVERVIEW POWER POWER POWER	format level area level component level format level format level area level component level format level area level component level
<b>Shared areas of interest</b>	YP 00 YP 00 YP 00 SD 50 SD 50 TKTB TKTB TKTB RC 90 RC 90	format level area level component level format level area level format level area level component level format level component level

*Areas of Interest used during VISA scoring, CASH alarm experiment*

<b>Scenario</b>	<b>Area of interest</b>
<b>1</b>	RD10 (high pressure heat exchanger), RD50 (top left corner superheater), RV00 RH10 (steam generator), Overview 1.
<b>2</b>	YA00 (lower thrid of screen ), YD00 (left bottom corner), Overview 1 (lower pump and parameters in left corner), Overview 2 (left bottom corner).

4	YA00 (middle of screen), YD00 (left half), YP00 (bottom third and top right corner), TKTE, YB00 (left middle of screen), Overview 1, Overview 2.
5	VCVA (seawater pumps, top screen), AX50 (middle of screen and top right corner), Overview 1 (pumps).
6	RC90 (top right corner), YA00 (pumps right part and bottom), YD00 (pumps), SD10 (valve s01/s03), YB00 (right bottom corner, s01, s02, s03), Overview 1, Overview 2.
8	TKTE (TK50 pump and left side of screen), YP00 (pumps in lower third of screen), TC00 (lower third of screen), YD00 (pumps in right half of screen), Overview 1, Overview 2 (right third of screen).
11	SK00, SD10, RD50, RQ00 (middle quadrant of bottom screen) Overview 1.
12	AX10, RM10, Overview 1.
13	RD10 (right top corner), SK00 (left top corner), SA50, TC00 (valve to sealing water, left bottom corner), YD00, Overview 1.
17	YB00, TKTE (TE23) RL00 (left bottom corner), Overview 1, Overview 2.
18	SP50, TKTE, Overview 1, Overview 2.
<b>General</b>	RA00P901, YA13P902, YP10L001, YA10T817.

In the HCA/HEAP study VISA was scored for scenario 3 and 4, thus, one template was made for each of the scenarios and one template containing AOIs was shared by both scenarios. The figure below shows the template made for scenario 3. The AOIs were grouped according to which format they belonged to. The template for scenario 3 had four formats, situated in the corners of the template. By pressing the blue area the corresponding format code was registered, by pressing the yellow area, the area code was registered, and finally, by pressing one of the images, or black areas, the component was registered. The area labelled long was used whenever the operator's gaze rested longer than one second at an AOI, i.e. when the gaze-line left the AOI.

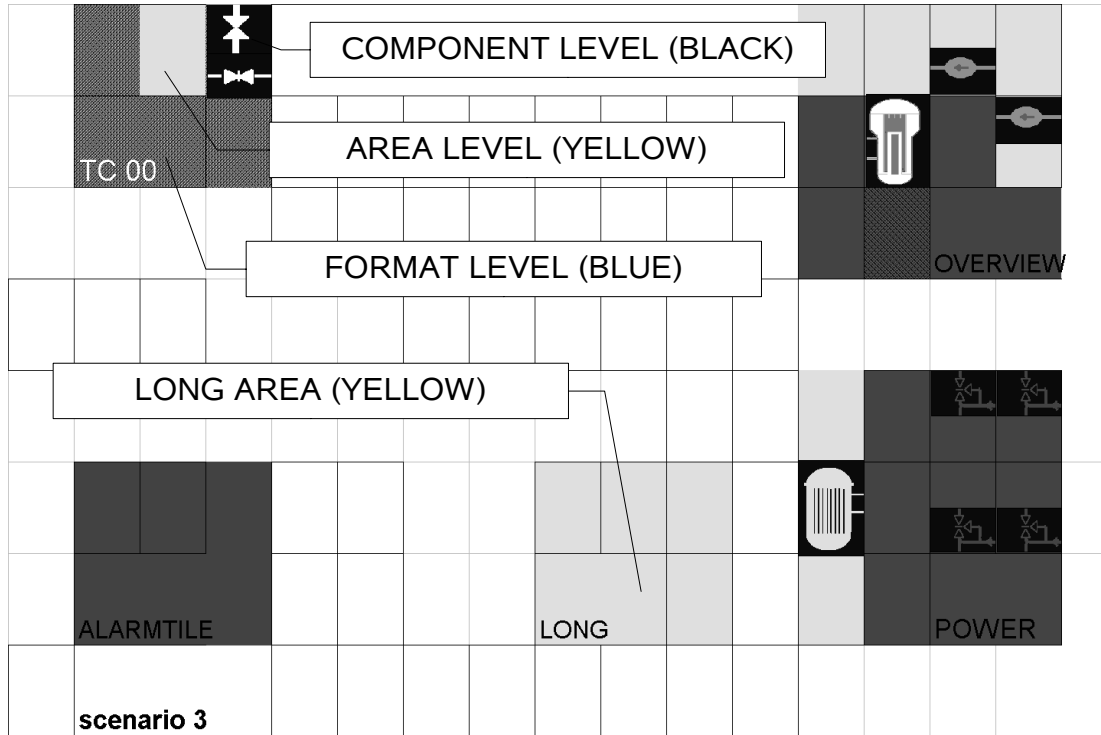


Figure: Template for scenario 3. The template is printed on foil in colour, and is placed on the touch sensitive tablet.

The EMT video was run at approximately 0.5 of actual speed, during the scoring. The scoring procedure was done twice for every subject. The specific scenario templates were used during the first run, and the general template, shared by both scenarios, was used in the second run.

The EyeCon software creates raw data logs in the format shown in the following table:  
 Excerpt of the EyeCon logging sheet. Operator 5, scenario 3, using the template shared by both scenarios. Together with the AOI-number, a time-code is also registered.

Time	AOI-Number	AOI-Name
00:26:37	3	YP EKSASKT
00:26:38	12	LONG
00:26:38	1	YP 00
00:26:40	12	LONG
00:26:41	3	YP EKSASKT
00:27:28	1	YP 00
00:27:37	3	YP EKSASKT
00:27:38	1	YP 00
00:27:38	2	YP OMRÅDE
00:27:41	6	TKTB
00:27:49	12	LONG
00:27:49	8	TKTB EKSAKT
00:27:52	8	TKTB EKSAKT
00:27:57	12	LONG
00:28:01	8	TKTB EKSAKT
00:28:03	12	LONG
00:28:07	6	TKTB
00:28:08	1	YP 00

Before any statistical analyses were performed, the logs were refined into a new format.

1. Every fixation was registered with a dwell time of 0.6 seconds, except whenever the Long area was pressed. In these instances, the amount of time passed since the last AOI was used instead.
2. For every subject, the time spent monitoring the different AOI's were summarised according to which level of AOI they were categorised: the format level included all AOIs; the area level included area and component AOIs; the component level only the component AOIs.
3. The period time varied between subjects, thus it was necessary to divide the three AOI level-scores with their respective period-times, resulting in three mean dwell time measures per minute.

## 10. APPENDIX II

Description of the procedure for scoring EDT

The EDT measure was based on data registered by the eye movement tracking (EMT) equipment. Several transformations were done to prepare data for statistical analysis. These changes were the following.

1. The packed data files produced by the eyetracker software, was unpacked and converted from binary to ASCII format by use of an uncompress program, giving the raw-data logs.
2. The raw-data logs contained 15 different variables. A Pearl script extracted 5 of these variables and made a<sup>1</sup>new file manageable for the program Eyetrack 2.
3. In Eyetrack 2 the mean fixation time was estimated. Period was specified in every individual case, as well as minimum fixation time and fixation area deviation. The period varied between subjects and scenarios, but the minimum fixation time and area deviation were constantly set at 150ms and 7 respectively. The table below shows the period time for all the subjects.
4. Every mean fixation time was read out of Eyetrack 2 and put into the data matrix.

---

<sup>1</sup> The Eyetrack 2 is a software package made for analysing EMT data, developed at Risø National Laboratory

*Time Windows used during EDT scoring in Eyetrack 2*

<b>Operator</b>	<b>Scenario</b>	<b>Period</b>	<b>Start</b>	<b>Stop</b>
1	3	1	01:00	15:58
1	3	2	24:19	29:27
2	4	1	01:00	16:23
2	4	2	28:14	37:14
3	4	1	01:00	16:05
3	4	2	23:02	31:51
4	3	1	01:00	16:48
4	3	2	22:41	31:50
5	3	1	01:00	16:05
5	3	2	25:47	34:45
6	4	1	01:00	16:05
6	4	2	20:34	29:31
7	4	1	01:00	16:09
7	4	2	21:42	30:39
8	3	1	01:00	16:05
8	3	2	21:41	30:49
9	3	1	01:00	09:06
9	3	2	14:18	22:22
10	4	1	01:00	16:02
10	4	2	26:40	36:19
11	4	1	01:00	16:04
11	4	2	23:22	32:20
12	3	1	01:00	16:52
12	3	2	22:52	32:16
13	3	1	01:00	16:02
13	3	2	26:25	35:27
14	4	1	01:00	16:06
14	4	2	22:52	31:50
15	4	1	01:00	16:05
15	4	2	22:25	31:25

## 11. CHOOSING MINIMUM FIXATION TIME AND FIXATION AREA DEVIATION

Analysis of the CASH alarm experiment, showed that a minimum fixation time set at 150ms and the area deviation parameter set at 5, was appropriate for the EDT measure. Similar results were found in the HEAP/HCA experiment. Which values are chosen for these parameters are not critical, as long as the values are kept within the range of 150-250ms for the minimum fixation time, and 4-7 for the area deviation measure. This is illustrated in the figure below which shows the values obtained for one subject, using four different values for both parameters.

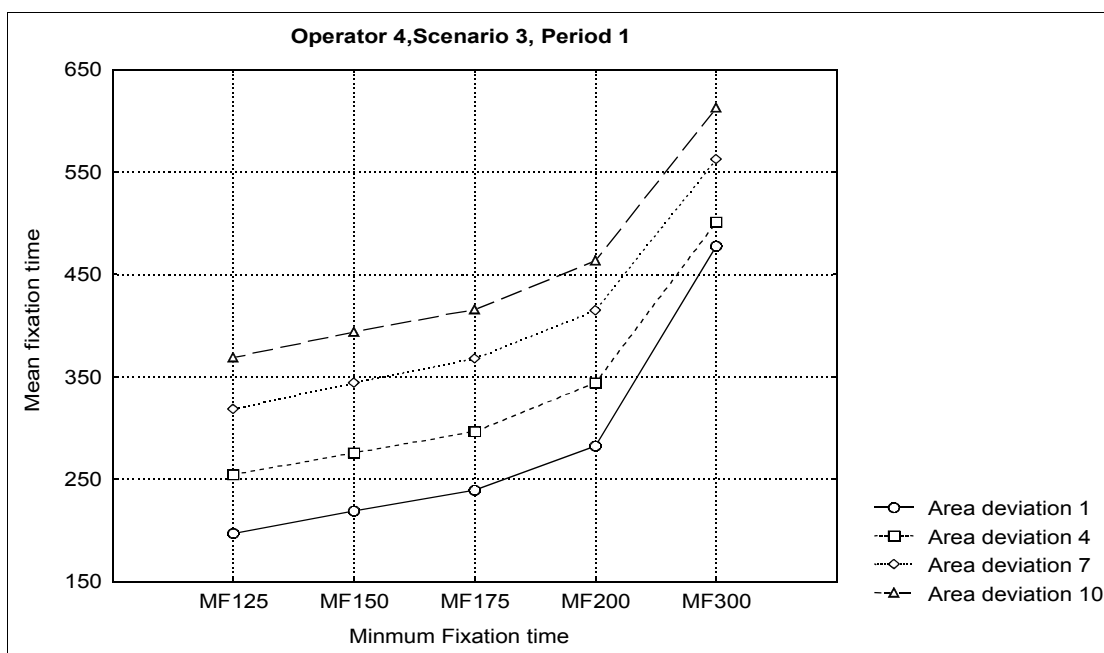


Figure: Line plot of mean dwell time obtained for operator 4, scenario 3, during period 1, using 4 different values for minimum fixation time and area deviation. .

## 12. APPENDIX III

Examples of categorisation of actions, used for scoring of Activity Level.

1	Navigation	"Window removed"
		"Menu button addressed"
		"Keyboard"
		"Incr"
		"Decr"
		"Controller"
		"Pressed"
2	Information gathering	"addressed"
		"Info on component"
		"Info on automatic"
3	Operations	"OPEN"
		"CLOSE"
		"MANUAL"
		"AUTO"
		"TRIP"
		"STAUP"
		"Confirmed reactor trip"
		"Scram"
		"Set new value"
4	Alarm interactions	"keyboard SIL"
		"kbd Global ACK"
		"kbd Global COM"
		"kbd Primary Global ACK"
		"kbd Primary Global COM"
		"kbd Secondary Global ACK"
		"kbd Secondary Global COM"
5	Format shift	"Select new picture"



## Paper G

Nystad, E., Drøivoldsmo, A. & Sebok, A. (2002). Use of radiation maps in a virtual training environment for NPP field operators. Proceedings of the man-machine systems research sessions at the enlarged Halden programme group meeting, Gol, Norway, 8th to 13 September 2002 (volume 2.) (pp. C5 P1. Halden, Norway: OECD Halden Reactor Project.



## **Extended Summary of Halden Work Report, HWR-681: Use of Radiation Maps in a Virtual Reality Training Environment for NPP Field Operators**

Espen Nystad, Asgeir Drøivoldsmo, Angelia Sebok

*OECD Halden Reactor Project*

*Phone: +47 69212200, Fax: +47 69 21 24 60,*

*E-mail:Espen.Nystad@hrp.no, Asgeir.Droivoldsmo@hrp.no, Angelia.Sebok@hrp.no*

VR may provide a useful tool for training because it offers the student direct experience with the material to be learned. In the nuclear industry, VR provides a safe method of training workers on tasks to be performed in radiation-exposed areas. VR training may be used for planning operations so they can be performed quickly and efficiently, thus reducing worker radiation doses. Finally, VR enables users to see things that are actually invisible, such as ionising radiation.

Evaluation of the efficiency of VR training is important to ensure that the skills or information learned in VR actually transfers to the real situation in which it is to be used. In cases where VR training is supposed to take the place of some other type of training, the old training method may be used as a benchmark with which to compare the VR training results.

Research investigating spatial learning using VR has not been able to give conclusive results regarding the effectiveness of VR training compared to map training or real-world training. While VR does provide a viable training technology, it requires more training time for the same level of proficiency compared to map or real-world training. Studies show conflicting results regarding whether VR, map, or real-world training provide the most effective training,

but these results depend heavily on the performance criteria used. Several VR training studies have shown that being active in the training situation leads to better performance than passively being presented the learning material, and there is evidence that a more restricted learning situation (guided training) reduces learning compared to a freer, less restricted learning situation (non-guided training).

This study evaluated three types of radiation distribution training for the Halden Boiling Water Reactor (HBWR) hall. Two training types were presented using VR models and computerised radiation maps; one free and one more restricted training type. The third type of training used a paper map showing the reactor hall and the radiation distribution in the hall.

The purposes of the study were to identify if:

- VR training will transfer to the actual working environment
- VR training is superior to map training for spatial knowledge training
- One type of VR training (guided or non-guided) is superior for spatial skills training

Participants were instructed to learn the radiation layout and practice walking along a route in the reactor hall using either a VR model or a map of the reactor hall. They were also instructed to learn to identify the location of points that they would later have to recognise and measure. The VR training included a guided condition, where users were shown the path they would need to learn, and a non-guided condition, where users were relatively free to explore and find the path that gave the least radiation. Following training, participants in the three training groups (five in each group) participated in the same experimental session. During the experimental session, participants applied the skills they had acquired in training by walking through the real reactor hall and performing the measuring tasks they had learned about in the training session. A number of performance measures were collected, including radiation awareness, presence, usability, the time used, subjectively reported transfer of training, and the number of measurement points correctly recognised.

Participants in the VR non-guided (active learning) condition demonstrated higher radiation awareness than the map condition. The VR non-guided condition also correctly recognised more measure points and reported higher transfer of training than the VR guided (passive learning) condition. The results of this study indicate that VR can provide a useful medium for training spatial skills. The results also indicate that the type of VR training is important, with active VR training resulting in better performance than passive VR training.

# Paper H

Nystad, E., Helgar, S. & Drøivoldsmo, A. (2002). Virtual reality verification of workplace design guidelines: A follow-up study. Proceedings of the man-machine systems research sessions at the enlarged Halden programme group meeting, Gol, Norway, 8th to 13 September 2002 (volume 2.) pp. C5 P9. Halden, Norway: OECD Halden Reactor Project.



# **Extended Summary of Halden Work Report, HWR-718: Virtual Reality Verification of Workplace Design Guidelines: A follow-up Study**

Espen Nystad, Stein Helgar, Asgeir Drøivoldsmo

*OECD Halden Reactor Project*

*Phone: +47 69212200, Fax: +47 69 21 24 60,*

*E-mail:Espen.Nystad@hrp.no, Stein.Helgar@hrp.no, Asgeir.Droivoldsmo@hrp.no*

## **1. BACKGROUND**

During the year 2000 a validation study of the Halden Reactor Project's utilisation of virtual reality (VR) technology in the process of verifying control room guidelines in human factors engineering design was conducted. It was found that the virtual mockup of the control room did not provide sufficient support for performing verifications of all the guidelines used in the study. A more thorough analysis of the results of the study revealed two major causes for the poorer performance in the virtual mockup as compared to the real control room. One was a low level of detail in the VR model, and the other was a lack of adequate measurement tools in the VR condition.

This report presents a follow-up study similar to the first study that aimed to validate the use of an improved VR application in the guideline verification process. Prior to the follow-up study, the areas that were found to be problematic in the first study were addressed and improvements were made to the VR application and the VR model of the control room. Using modelled operating panels instead of textures gave a higher level of detail. Tools for measuring fonts, angles and illumination were included and the navigation was completely redesigned. As in the first study, the objective of this validation study was to determine how accurately the control room could be verified according to guidelines in a virtual control room compared to a physical mockup of a control room.

## **2. METHOD**

Each of twelve crews performed a guideline verification of a control room in two conditions: the real condition and the VR condition. In the real condition subjects performed measurements and evaluations in the actual Halden Boiling Water Reactor control room using appropriate measurement tools. In the VR condition the subjects

used version 1.0 of IFE's CREATE Verification Tool that presented the VR model of the control room and the integrated virtual measurement tools. In both conditions the guidelines were presented with the Verification Tool and the control room was rated in one of four categories according to how well it conformed to the guideline. The design was counter balanced by letting half of the crews perform the real control room guideline verification first, while the other half performed the VR model verification first. 18 subjects from the Halden Boiling Water Reactor participated in the study. The six control room and field operators were each paired with one of the subjects from the MTO sector and acted as subject-matter experts during the verifications. The six remaining subjects performed verifications individually.

The sections of NUREG-0700 (1996) that were used in the study were Consoles, Control room configuration, Illumination, Panel layout, and Panel labelling. The data collected were ratings for each of the 50 guidelines, the time used to carry out the verification for each guideline and it also included the subjects' comments and snapshots for each guideline. At the end of the verification session in the VR condition, the subjects evaluated the usability of the VR application.

### **3. RESULTS AND CONCLUSIONS**

The data were analysed to find the amount of agreement between the subjects in the ratings for each guideline. This was used as a measure of the reliability of the verifications. The number of ratings where the control room could not be verified was counted. Differences between the virtual and the real control room in this respect were an indication of problematic aspects of the VR tools or VR model.

The results showed some usability problems in the font tool and the distance measurement tool, although the distance measurements in VR were as reliable as the measurements made in the real control room. It was difficult to perform verifications in the VR model for guidelines related to shadowing, glare and ambient illumination, but the verification of direct illumination worked well. The detail of the instrument panels and the rest of the control room seem to be sufficient.

### **4. REFERENCES**

NUREG-0700. (1996). *U.S. Nuclear Regulatory Commission. Human-System Interface Design Review Guideline* (NUREG-0700, Rev. 1, Vol. 1). Washington, DC, USA: The NRC Public Document Room.



# Paper I

Drøivoldsmo, A. (2002). The Halden Questionnaire help pages [On-line]. Available:  
<http://psyweb.svt.ntnu.no/asgeird>: Internet.



# Halden Questionnaire



Asgeir Drøivoldsmo

## Introduction

Reliable and cost-effective data collection is critical for quantitative research. Administration and scoring of questionnaires in an experiment or a survey data collection are time-consuming and error-prone tasks. Replacing pen and paper based questionnaires with a computer-based system gives a more reliable test outcome and opens a number of new ways of test presentation.

The Halden Questionnaire is a user-friendly tool for time-efficient management of questionnaire data. The system is an entirely web-based package allowing you to perform all the necessary steps from questionnaire design to preparing data for analysis.

All the necessary steps for creating questionnaires, experimental setup, data collection and preparations for analysis can be done from a web browser without the need for additional software. The system is still a prototype, but it has demonstrated speed and reliability in a number of experiments at the Halden Reactor Project.

The Halden Questionnaire can be installed and run on any type of computer from a mainframe Unix machine to a laptop PC or Mac. Depending on the purpose of the study, a single computer, intranet, or Internet solution can be used. User access is controlled at multiple levels in a way that makes it possible to share projects between several researchers who design and run the project together.

## What can I use the system for?

The system is designed for all sorts of questionnaire data collection using the Internet. The system has been tested with targeted email messages, performance measurement in experimental studies, usability testing and real world field experiments. All necessary steps from design to the production of the final data matrix have support in a self-explanatory main menu (Figure 1).

## Experiments and tests

The Halden questionnaire was originally made as a tool for saving resources in experimental data gathering and usability testing. The system has a built-in module for experimental runs where the independent variables in the experimental design are inserted in an “experimenter’s test page” for easy administration to test subjects. A dedicated results module is made for preparation of data in order to export it to a statistical software package. This module also contains an easy-to-use function for testing data against a teacher’s copy.

## Surveys

Surveys can be divided into two categories: one category where the respondent's name is known, and one anonymous category. The Halden Questionnaire has currently only implemented functionality for surveys where the respondent is provided with a username and password. Functionality covering anonymous surveys is to be implemented in the next version.

## Using the system

As shown in Figure 1, designing and running a data collection project is accomplished by following the steps from 1 to 5. The following introduction will show the main features under each step needed to perform a project.

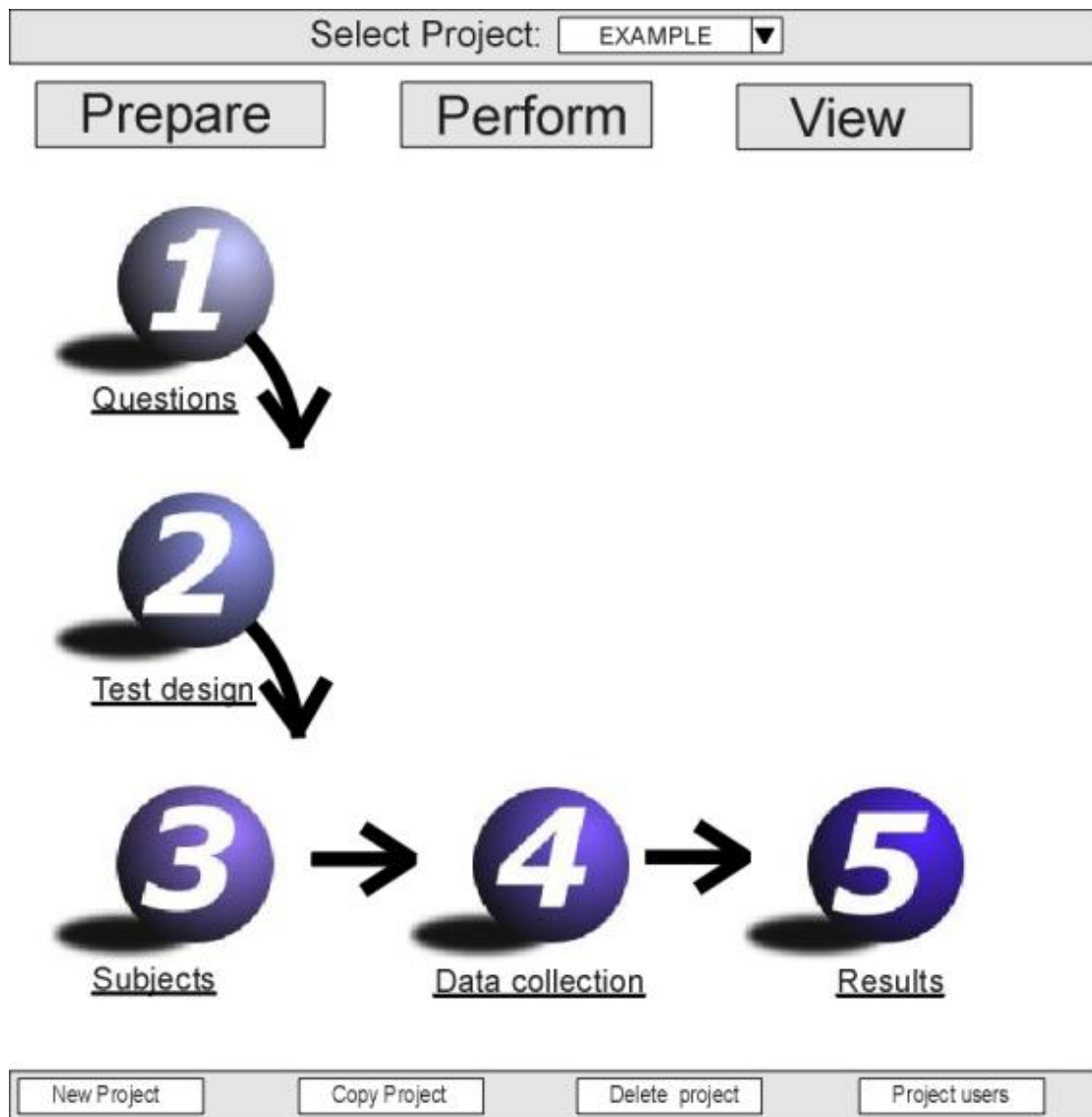


Figure 1. Halden questionnaire main page NB! Copy project is not implemented yet.

## Getting Internet access to the system

If the system is running on an Internet server, and a new user wants to start a project, the first thing to do is to send an email to the responsible system maintenance manager (a mailto button is found on the system entry page). The maintenance manager will then provide a username and a password for the researcher, necessary to enter and start the work. If the system runs on a local computer, the researcher can then control all passwords and access data using the user set-up and administration tool.

## Add, copy, delete and share projects

From the menu bar at the bottom of the main page (Figure 2), a project can be created, copied or deleted. Add a new 'project' by clicking the "New project" button or select an already created project from the drop-down menu at the top of the page and click the Switch button on the top menu line.



Figure 2. Project administration functions are found on the bottom main page menu.

The creator of a project is the owner of the project, and can share the project with colleagues using the "Project Users" function. Looking at Figure 3, sharing a project is done by choosing a name from the dropdown menu. If a new column is needed, the "Add user" button is pressed. The "Finish" button should always be pressed after choosing one or more names from the pull-down menus.

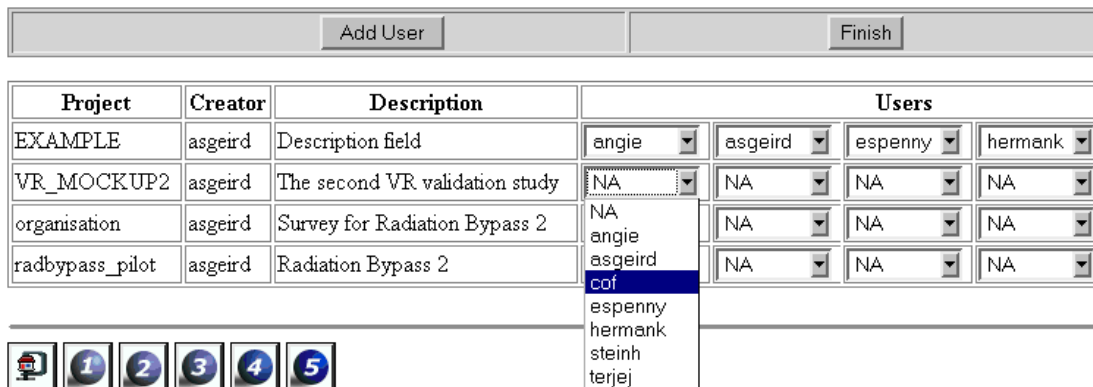


Figure 3. From the "Project Users" menu, projects can be shared with other users.



## Questions

After creating a new project, the first step is to create the questionnaires. Clicking ball one in the main menu gives the page for creating new questionnaires (Figure 4). This page shows a listing of already created questionnaire pages. For a new project, this list will be empty.

<input type="button" value="New Page"/>			
Page	Date	User	Info
<a href="#">1</a>	<a href="#">Sun Jun 9 21:37:32 MDT 2002</a>	<a href="#">asgeird</a>	<a href="#">Likert scale 7 items</a>
<a href="#">2</a>	<a href="#">Sun Jun 9 21:44:00 MDT 2002</a>	<a href="#">asgeird</a>	<a href="#">Forced choice</a>
<a href="#">3</a>	<a href="#">Sun Jun 9 21:50:58 MDT 2002</a>	<a href="#">asgeird</a>	<a href="#">Semantic differential</a>
<a href="#">4</a>	<a href="#">Sun Jun 9 22:10:04 MDT 2002</a>	<a href="#">asgeird</a>	<a href="#">Attributional style</a>
<a href="#">5</a>	<a href="#">Sun Jun 9 22:23:22 MDT 2002</a>	<a href="#">asgeird</a>	<a href="#">Tolerance for design approaches</a>
<a href="#">6</a>	<a href="#">Mon Jun 17 22:30:20 2002</a>	<a href="#">livhb</a>	<a href="#">Using pictures in the question</a>
<a href="#">7</a>	<a href="#">Tue Jun 25 9:06:50 2002</a>	<a href="#">asgeird</a>	<a href="#">test without slash</a>

A total of 7 steps

Figure 4. Create new page or edit an existing page

Select one of the existing pages to edit or add a question (Figure 4). A page can be the entire questionnaire or just one part (page) of the questionnaire. Click the 'New Page' button to add a new page.

<input type="button" value="Next"/>		<input type="button" value="Clear"/>
Num of Questions	Type of Question	Type of Box
<input type="text" value="1"/>	<input type="text" value="7-Options"/>	<input type="text" value="Radio"/>
<hr/>		
<b>Name of New Page</b>		
<input type="text" value="Likert Scale 7 items"/>		

Figure 4. Create new page

Select 'Num of Questions', 'Type of Questions', 'Type of Box' and fill in the field for 'Name of New Page', then click 'Next'. This takes you to Figure 5. Fill in the adequate labels, select the positioning of the labels and click 'Finish'.

In the 'Num of Questions' pull-down menu the number of questions or objects (1-10) can be selected. If you plan to have more than 10 questions on a page, you should start with 10 and then select the same page again for adding the remaining questions.

The different choices in the 'Type of Question' pull-down are presented in Table 1.

Name	Type	Description
2-Options	Radio button or checkbox	Radio-buttons are for single responses; checkboxes are for multiple responses
:		
:		
11-Options		
Short Answer	Text input field	Default size of the short answer text input field is 5 characters
Long Answer	Text input field	Default size of the long answer text input field is 55 characters
Multi-Media	All types of multi media elements like pictures, sound, video, vr-models, flash animations, etc.	
Hyperlink	URL pointing to a resource which can be handled by a www-browser	Useful for referring to additional pages not generated in the questionnaire
Heading1	Huge heading	Test or question heading
Heading2	Medium heading	Test or question heading
Heading3	Small heading	Test or question heading
Normal	Normal size text	Text input field for writing instructions and questions that should not have radio-buttons or checkboxes on the same line in the questionnaire.

*Tip*

Since the finished questionnaire displays headings and instructions, it is necessary to think through how the page should look before you add the questions. The Halden Questionnaire does not contain a point-and-click editor and therefore the sequence of entering the different elements in a page is important. Adding a heading after generating the questionnaire page is a difficult, but not impossible, job. Figure 5 shows the step of creating the labels for the questions.

Finish										Clear
Radio Box Question, Label	End L	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7	End R	
Above	Not a	1	2	3	4	5				

Figure 5. Create labels page

Figure 5 shows the result of choosing ‘heading3’ (the small heading) for the question header, normal text for instructions, ‘7-option’ radio buttons with the headings placed above, and filling in the end labels.

Next			Clear
Num of Questions	Type of Question	Type of Box	
1	2-Options	Radio	

---

<b>Questions in Page: Likert scale 7 items</b>								
<b>How self reliant do you find yourself?</b>								
Answer code: Select the appropriate number for each trait:								
Not at all self-reliant	1	2	3	4	5	6	7	Very self-reliant
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Figure 6. How the ‘Page’ now looks

To make changes, click the Edit button to open the code page (Figure 7). Make the changes in the code. Each entry in the table represents one question. To delete a row, unmark the desired checkboxes. Press ‘Finish’ to save.



To add a new question simply repeat the steps from Figure 4; Select 'Num of Questions', 'Type of Questions', 'Type of Box' and 'Name of New Page', then click 'Next' again.

<input type="button" value="Finish"/>		<input type="button" value="Clear"/>
---------------------------------------	--	--------------------------------------

Date	User	Name
<input type="text" value="Sun Jun 9 21:37:32 MDT 2002"/>	<input type="text" value="asgeird"/>	<input type="text" value="Likert scale 7 items"/>

Questions	
<input checked="" type="checkbox"/>	<pre>&lt;TR&gt;&lt;TD WIDTH="100%"&gt;&lt;H3&gt;How self reliant do you find yourself?&lt;/h3&gt; &lt;/td&gt;</pre>
<input checked="" type="checkbox"/>	<pre>&lt;TR&gt;&lt;TD WIDTH="100%"&gt;Answer code: Select the appropriate number for each trait:&lt;/td&gt;</pre>
<input checked="" type="checkbox"/>	<pre>&lt;TR&gt;&lt;TD&gt;&lt;TABLE width="100%" BORDER="0" ID="Tab1"&gt;&lt;TR&gt;&lt;TH WIDTH="11%"&gt; Not at all self -reliant&lt;/th&gt;&lt;TH WIDTH="11%"&gt;1&lt;br&gt;&lt;INPUT TYPE="radio" name="que_1" value="1"&gt;&lt;/th&gt;&lt;TH WIDTH="11%"&gt;2&lt;br&gt;&lt;INPUT TYPE="radio" name="que_1" value="2"&gt;&lt;/th&gt;&lt;TH WIDTH="11%"&gt;3&lt;br&gt;&lt;INPUT</pre>

Figure 7. Edit code page



## Test design

Test design is the administration plan for tests/questionnaires; each test is an individual questionnaire or a set of questionnaires.

<input type="button" value="Add Col"/>		<input type="button" value="Add Row"/>		<input type="button" value="Finish"/>	
Test 1	Likert scale 7 items	Attributional style			
Test 2	Forced choice	NA			

*Figure 8. The example test design page shows two tests: test 1 - which includes the questionnaires "Likert scale 7 items" and "Attributional style," and test 2 - which includes only the "Forced choice" questionnaire.*

Figure 8 shows the page where the sequence of questionnaires to be given to the subject is set. Click 'Add Row' to add a new test and 'Add Col' to add a new questionnaire in the test. The questionnaires are picked from the pull-down menu in each table entry. NB! It is important to select all tests in a row before adding a column and vice versa.

If you want to delete a row or column, enter NAs only in the row or column you want to remove, then push "Finish." This will delete the row or column.



## Run plan

Add Col		Add Row				Finish		
Date	Name	Cond 0	Cond 1	Cond 2	Cond 3	Cond 4	Tests	
12.01	su01	high	No-shocl	first			1	NA
14.02	su02	low	1200V	first			1	NA
14.02	su03	high	No-shocl	first			1	NA
14.02	su04	high	1200V	second			1	2
14.02	su05	low	No-shocl	second			1	2
14.02	su06	low	1200V	second			1	2
xxx	asgeird	xxx	xxx	xxx			1	2

Figure 9. Run plan

Figure 9 shows the run plan set-up. Click 'Add Row' to enter test subjects and 'Add Col' to determine which tests each subject is going to take. Usernames (test subjects) are defined in a separate file. 'Cond' fields can be used for additional information, e.g. education level or nationality. NA is none. Click Finish to save. In Figure 9 subject su01, su02 and su03 are set up to run only test1. Su04, su05 and su06 are running first test1 and then test2. To test the order of questionnaires and test before the "real" data collection starts, the designer should always put her/him self up as the last subject in the test. Further advice for this testing is found in the next section.

The "Date" and "Cond" fields are text entry fields. Since the contents of these fields are transferred later to the final data matrix, it is easier to specify the experimental design at the same time as the run plan is made. Having the experimental design transferred to the data matrix makes it easier to overview the data during the data collection and saves the job of entering it in the statistical software used for the analysis.

## Data collection

In the data collection module there are two different views of data collection. The view presented in Figure 10 is the designer's view for testing the questionnaire or running experiments. The purpose of this view is to provide an efficient method of test distribution. When running an experiment the experimental leader should log in with his own name (s(he) must be a member of the project) and then chose the "Start" entry for the subject that is supposed to run a test.

(NB! not yet implemented). This page also gives the designer a shortcut to the test and the chance to return back into the design modules after testing. To step through the tests click, on one of the text labels under "Tests". When finished, push the "Return to designer view" button. To run the test the same way as the subjects, click the "Start" label to start the test wizard. After entering the test wizard, no return route to the design modules is provided.

Date	Name	Cond 0	Cond 1	Cond 2	Cond 3	Cond 4	Wizzard	Tests
12.01	su01	high	No-shock	first			<a href="#">Start</a>	<a href="#">test1</a>
14.02	su02	low	1200V	first			<a href="#">Start</a>	<a href="#">test1</a>
14.02	su03	high	No-shock	first			<a href="#">Start</a>	<a href="#">test1</a>
14.02	su04	high	1200V	second			<a href="#">Start</a>	<a href="#">test1</a> <a href="#">test2</a>
14.02	su05	low	No-shock	second			<a href="#">Start</a>	<a href="#">test1</a> <a href="#">test2</a>
14.02	su06	low	1200V	second			<a href="#">Start</a>	<a href="#">test1</a> <a href="#">test2</a>
xxx	asgeird	xxx	xxx	xxx			<a href="#">Start</a>	<a href="#">test1</a> <a href="#">test2</a>

Figure 10. Data collection page for designer testing of the questionnaire

After clicking on the "Start" link, the system prompts for a password. The password for su01 is su01 and su02 has password su02, etc.

In the example in Figure 11, su04 is going to fill in the two tests created in this help manual. Each test includes different steps (set of questionnaires, i.e., Figure 8). The green backgrounds under "test1" and "test2" indicate that no data is registered for su04.

*Tip!*

The default picture presented in Figure 11 can be replaced with a front page for the specific test. This is done in the Questions module by building a question by the same procedure as described under 'Questions' in section 1. To make it the test front page give it the same name as the current project, i.e., EXAMPLE in this example.

# HALDEN VIRTUAL REALITY CENTR

Applying advanced visualisation technologies and human factors to solve real-world challe

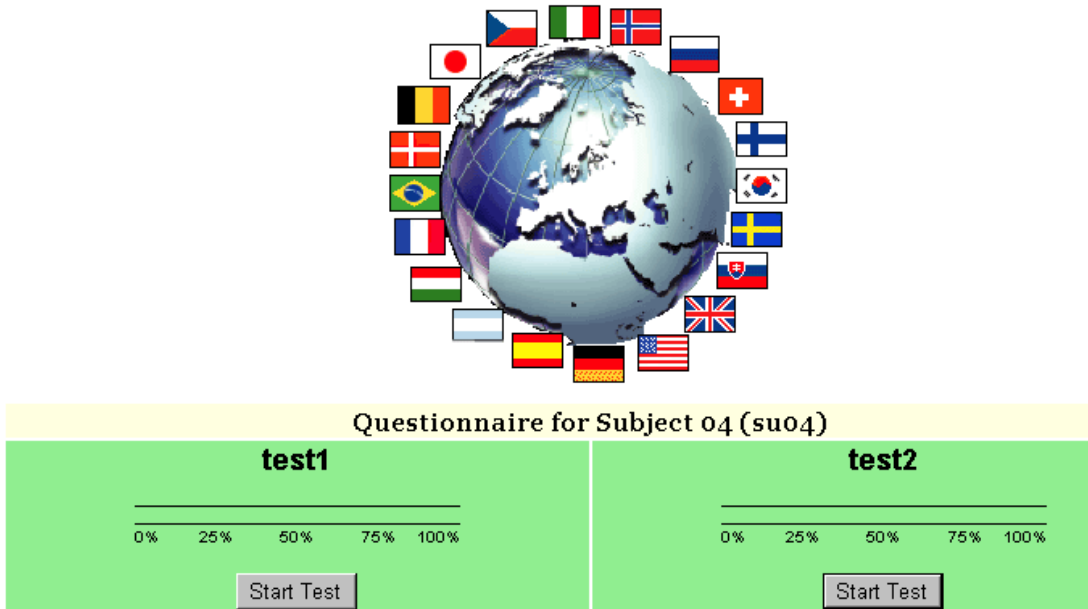


Figure 11. The test wizard page.

After the subject has finished the first test, the background under test1 turns red (see Figure 12) to indicate that data has been collected for this test. If for some reason anything should go wrong during the filling in of a test, the system will ask whether to continue or restart the test when the web browser has been restarted.

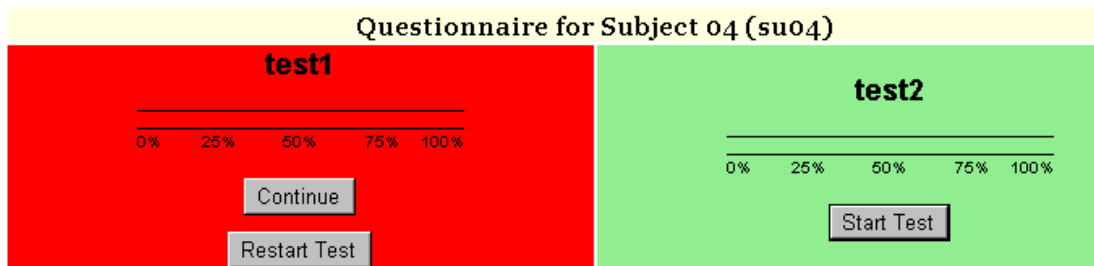


Figure 12. The test wizard page where su04 has partly finished test1 and when logging in to the system again, the system prompts for "continue" or "restart" of the test.



## Results

Figure 13 shows an overview page of the results. The "Numerical", "Text" (not implemented), and "Time Stamp" check boxes decide what will be seen in the raw data report.

Raw Data Report									
<input type="button" value="Submit"/>									
<input checked="" type="checkbox"/> Numerical			<input type="checkbox"/> Text			<input type="checkbox"/> Time Stamp			

Additional Reports									
<input type="button" value="Text in Readable Format"/>			<input type="button" value="Teachers Copy"/>			<input type="button" value="Export to Text (ASCII)"/>			

Name	Date	Cond1	Cond2	Cond3	Cond4	Cond5	test1			test2
su01	12.01	high	No-shock	first			1	2		
su02	14.02	low	1200V	first			-9999	-9999	-9999	-9999
su03	14.02	high	No-shock	first			-9999	-9999	-9999	-9999
su04	14.02	high	1200V	second			1	I got the sack	1	1
su05	14.02	low	No-shock	second			1			-9999
su06	14.02	low	1200V	second			-9999	-9999	-9999	-9999
asgeird	xxx	xxx	xxx	xxx			1	Need some change	1	1

Figure 13. Example of a results page

The "Text in Readable Format" button will produce a text report of all free text input fields in the questionnaires.

Compare Teachers Copy to Different Run(s)								
Compare		<a href="#">Toggle selections</a>						
Teachers Copy	Name	Date	Condo	Cond1	Cond2	Cond3	Cond4	Compare To
<input type="radio"/>	12.01	su01	high	No-shock	first			<input checked="" type="checkbox"/>
<input type="radio"/>	14.02	su02	low	1200V	first			<input checked="" type="checkbox"/>
<input type="radio"/>	14.02	su03	high	No-shock	first			<input checked="" type="checkbox"/>
<input type="radio"/>	14.02	su04	high	1200V	second			<input checked="" type="checkbox"/>
<input type="radio"/>	14.02	su05	low	No-shock	second			<input checked="" type="checkbox"/>
<input type="radio"/>	14.02	su06	low	1200V	second			<input checked="" type="checkbox"/>
<input checked="" type="radio"/>	xxx	asgeird	xxx	xxx	xxx			<input type="checkbox"/>

Figure 14. Example using the teacher's copy function

The teachers' copy function makes use of one of the test subjects as a key to correct answers selected from the "Teachers copy" column to the left in Figure 14. The answers from this subject are then compared to all the subjects selected in the rightmost column in the figure.

What is seen in the raw data report will be exported to the results files when "Export to ASCII" is chosen. The results are saved in *report.txt*, which can easily be read by statistical analysis software or spreadsheet systems. The default delimiter is the semicolon (ask the system manager to change this if necessary).

## **Final comment**

This system has been made for the purpose of effective:

- questionnaire design
- experimental setup
- data collection
- preparation of data for analysis

The system runs on Unix, Windows, Linux and MacOSX on both the server and client side, requiring the Apache web-server and perl5 (see the technical section for further details about installation and system requirements).